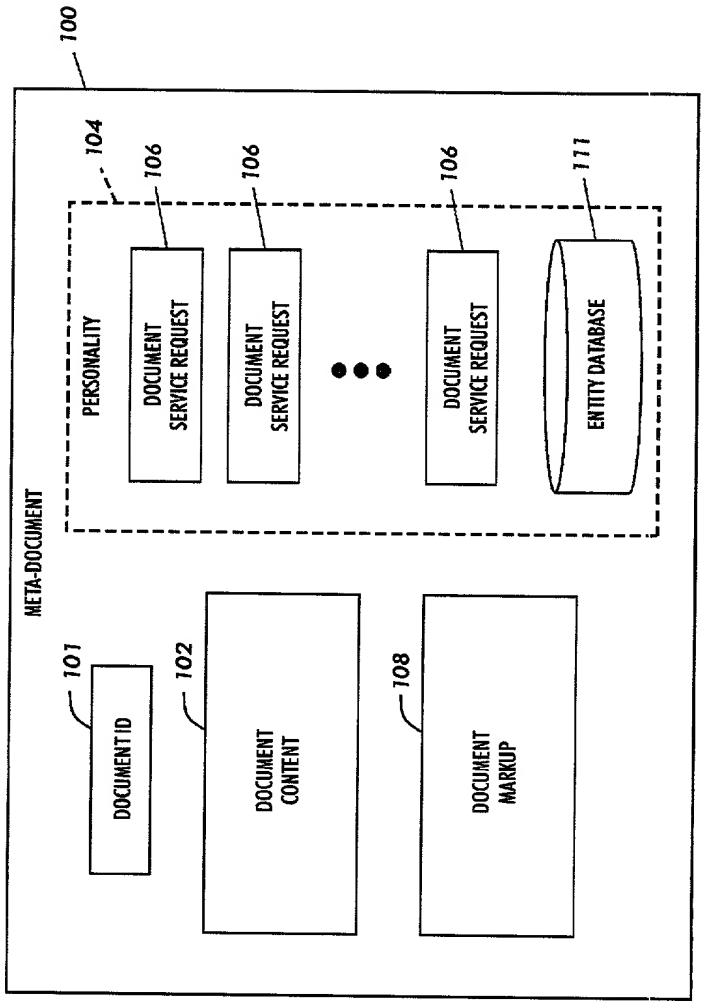
100
META-DOCUMENT
104
PERSONALITY
106
DOCUMENT SERVICE REQUEST
106
DOCUMENT SERVICE REQUEST
106
DOCUMENT SERVICE REQUEST
111
ENTITY DATABASE
101
DOCUMENT ID
102
DOCUMENT CONTENT
108
DOCUMENT MARKUP

FIG. 1

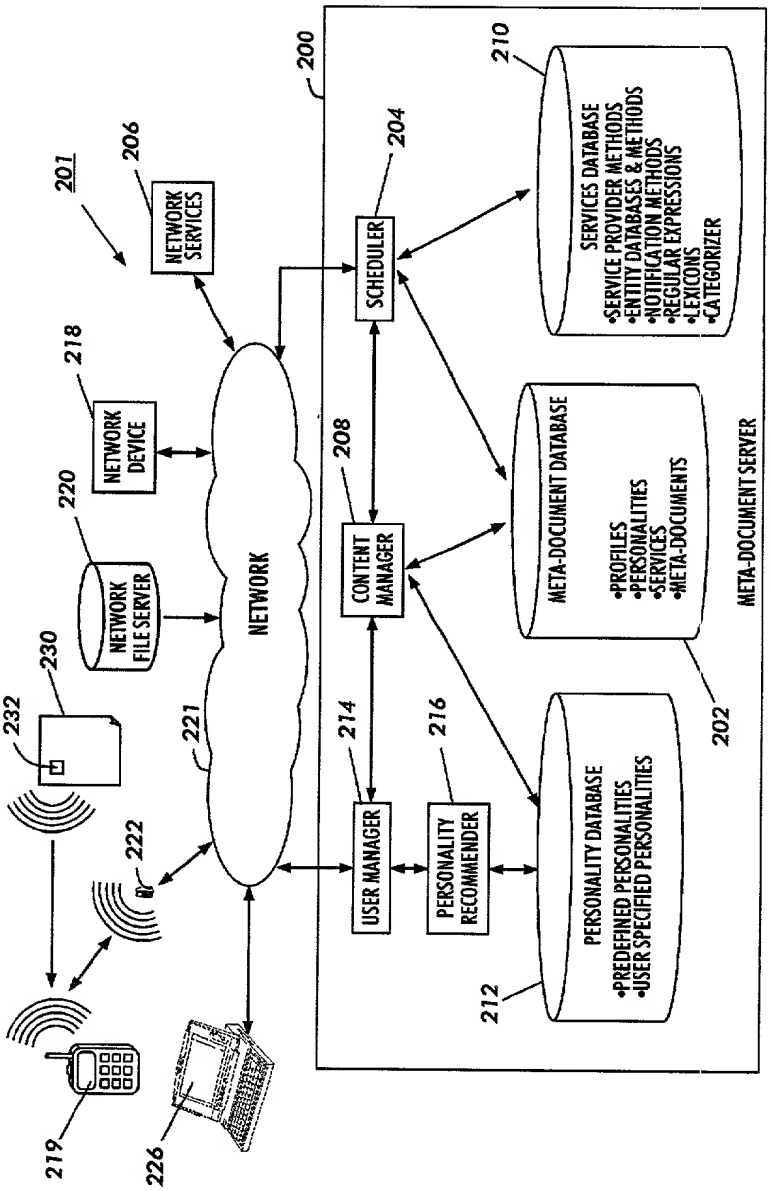
200
201
206
NETWORK SERVICES
218
NETWORK DEVICE
220
NETWORK FILE SERVER
230
232
221
NETWORK
222
219
226
204
SCHEDULER
208
CONTENT MANAGER
214
USER MANAGER
216
PERSONALITY RECOMMENDER
210
SERVICES DATABASE
•SERVICE PROVIDER METHODS
•ENTITY DATABASES & METHODS
•NOTIFICATION METHODS
•REGULAR EXPRESSIONS
•LEXICONS
•CATEGORIZER
META-DOCUMENT DATABASE
•PROFILES
•PERSONALITIES
•SERVICES
•META-DOCUMENTS
202
PERSONALITY DATABASE
•PREDEFINED PERSONALITIES
•USER SPECIFIED PERSONALITIES
212
META-DOCUMENT SERVER

FIG. 2

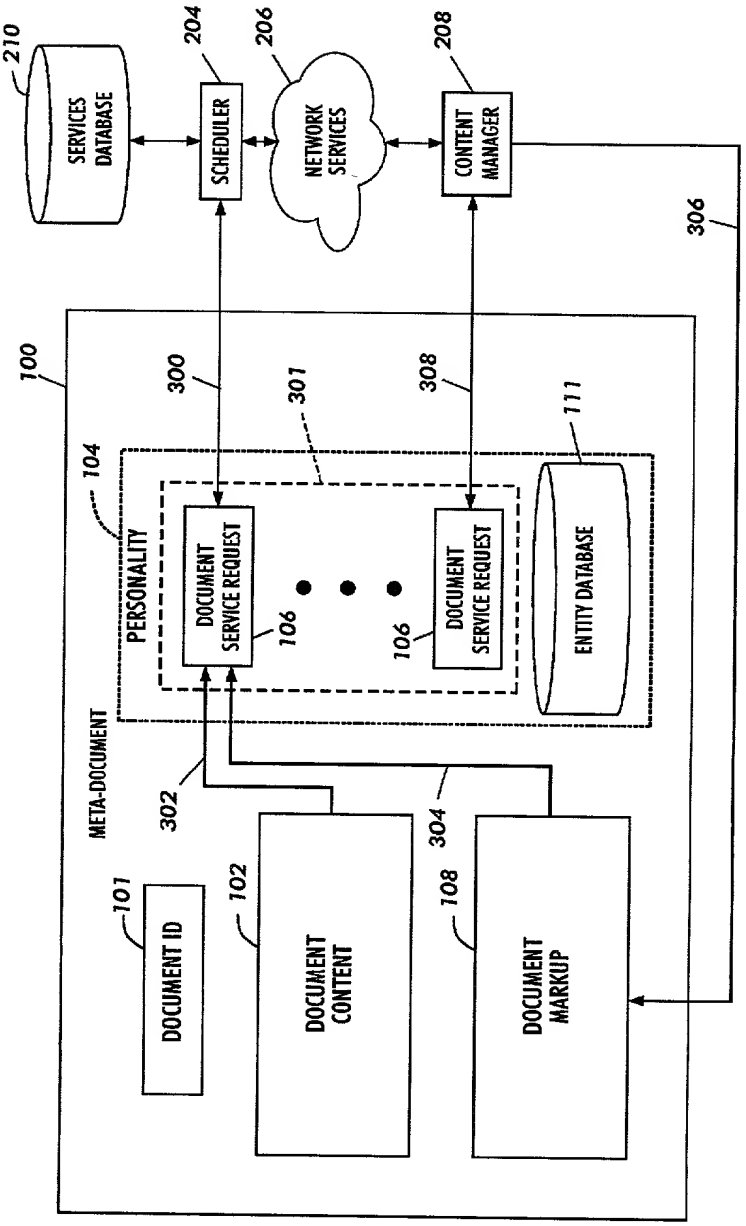
210
SERVICES DATABASE
204
SCHEDULER
206
NETWORK SERVICES
208
CONTENT MANAGER
306
100
300
301
308
111
104
PERSONALITY
DOCUMENT SERVICE REQUEST
106
106
DOCUMENT SERVICE REQUEST
ENTITY DATABASE
META-DOCUMENT
302
304
101
DOCUMENT ID
102
DOCUMENT CONTENT
108
DOCUMENT MARKUP

FIG. 3

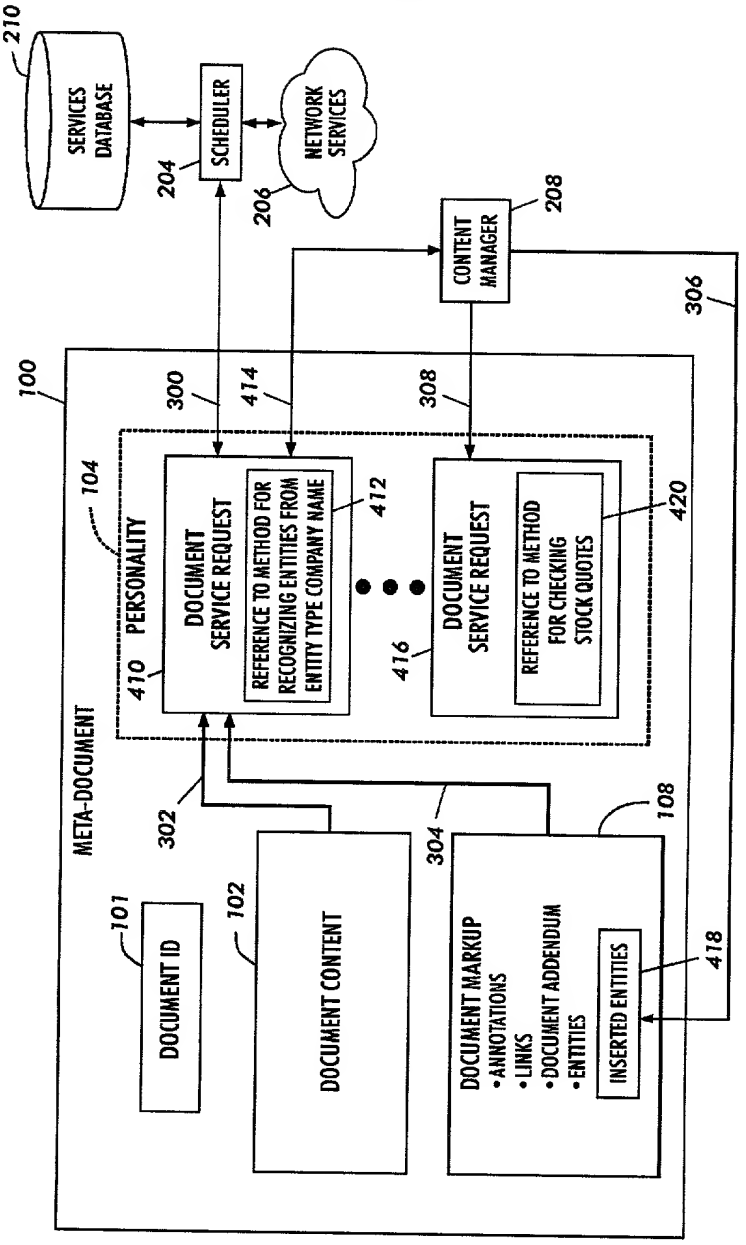
210
SERVICES DATABASE
SCHEDULER
204
NETWORK SERVICES
206
CONTENT MANAGER
208
306
100
META-DOCUMENT
300
414
308
104
PERSONALITY
410
DOCUMENT SERVICE REQUEST
REFERENCE TO METHOD FOR RECOGNIZING ENTITIES FROM ENTITY TYPE COMPANY NAME
412
416
DOCUMENT SERVICE REQUEST
REFERENCE TO METHOD FOR CHECKING STOCK QUOTES
420
302
304
108
101
DOCUMENT ID
102
DOCUMENT CONTENT
DOCUMENT MARKUP
• ANNOTATIONS
• LINKS
• DOCUMENT ADDENDUM
• ENTITIES
INSERTED ENTITIES
418

FIG. 4

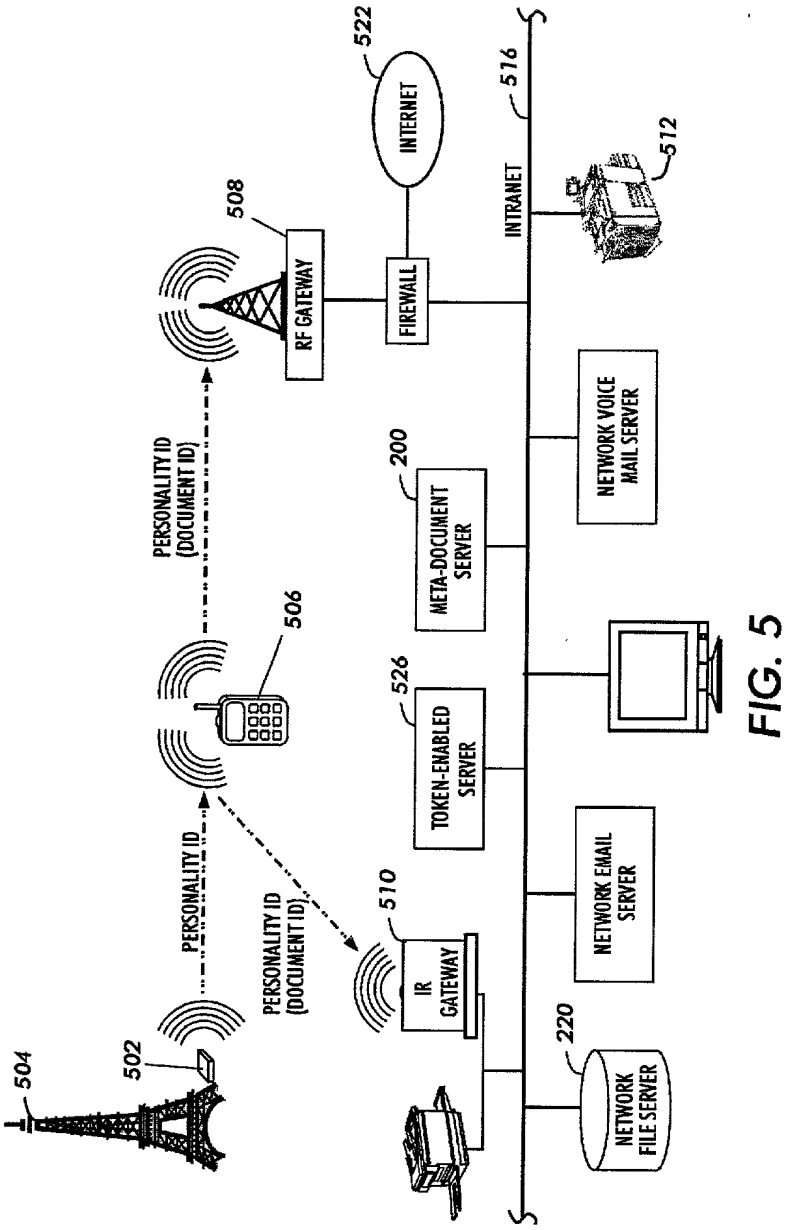
INTERNET
522
INTRANET
516
512
508
RF GATEWAY
FIREWALL
PERSONALITY ID (DOCUMENT ID)
NETWORK VOICE MAIL SERVER
200
META-DOCUMENT SERVER
506
526
TOKEN-ENABLED SERVER
PERSONALITY ID
PERSONALITY ID (DOCUMENT ID)
510
IR GATEWAY
NETWORK EMAIL SERVER
504
502
220
NETWORK FILE SERVER

FIG. 5

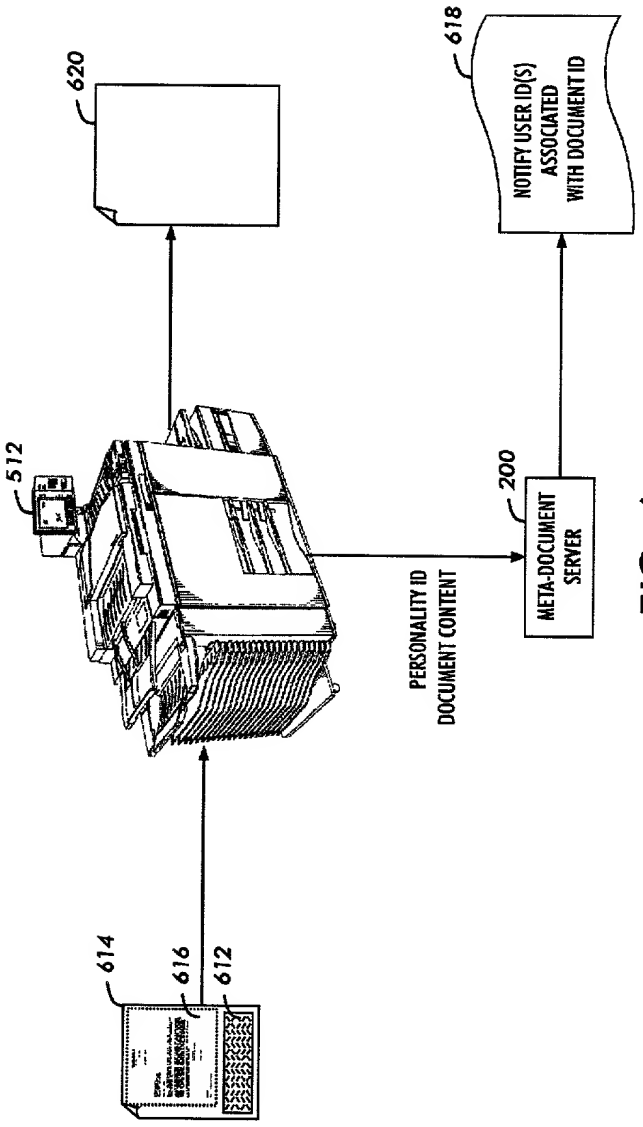
620
618
NOTIFY USER ID(S) ASSOCIATED WITH DOCUMENT ID
512
200
META-DOCUMENT SERVER
PERSONALITY ID
DOCUMENT CONTENT
614
616
612

FIG. 6

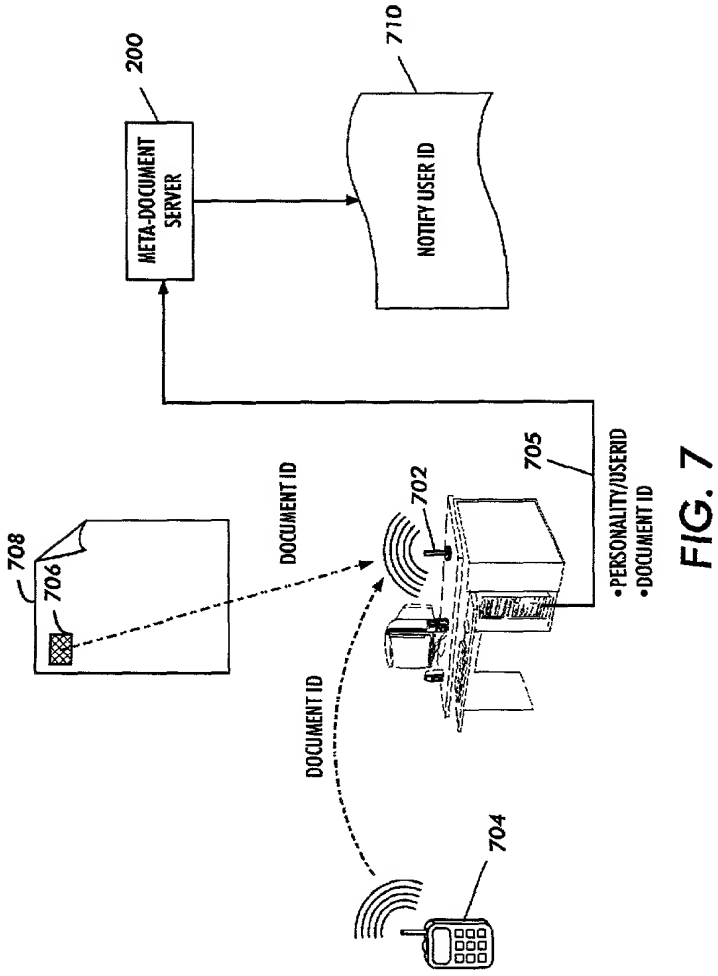
200
META-DOCUMENT SERVER
710
NOTIFY USER ID
705
•PERSONALITY/USERID
•DOCUMENT ID
702
DOCUMENT ID
708
706
DOCUMENT ID
704

FIG. 7

8/69

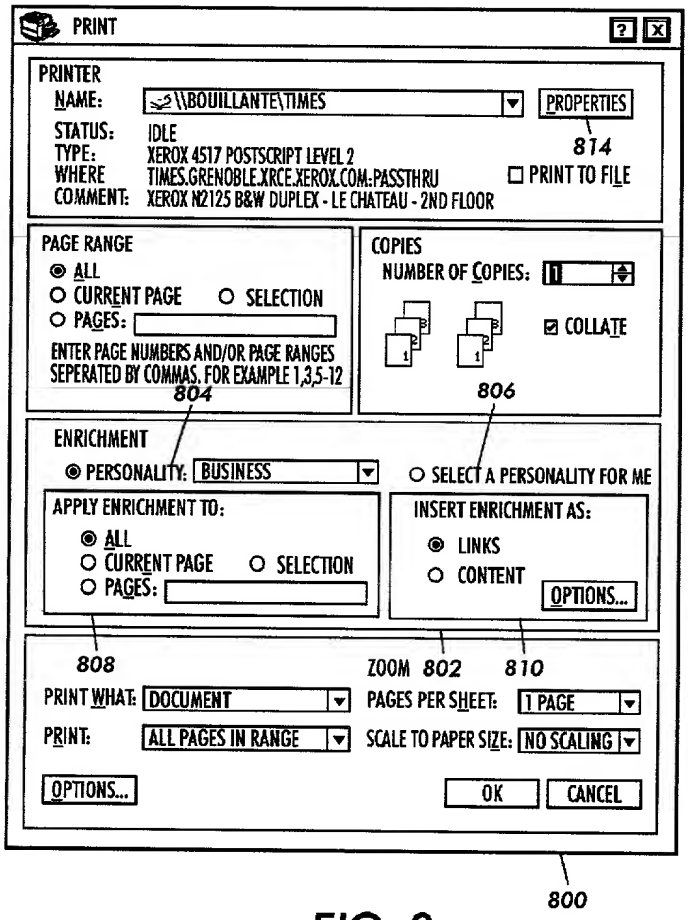
PRINT
PRINTER
NAME: \\BOUILLANTE\TIMES
PROPERTIES
814
STATUS: IDLE
TYPE: XEROX 4517 POSTSCRIPT LEVEL 2
WHERE TIMES.GRENOBLE.XRCE.XEROX.COM:PASSTHRU
COMMENT: XEROX N2125 B&W DUPLEX - LE CHATEAU - 2ND FLOOR
PRINT TO FILE
PAGE RANGE
ALL
CURRENT PAGE
SELECTION
PAGES:
ENTER PAGE NUMBERS AND/OR PAGE RANGES SEPERATED BY COMMAS. FOR EXAMPLE 1,3,5-12
804
COPIES
NUMBER OF COPIES: 1
COLLATE
806
ENRICHMENT
PERSONALITY: BUSINESS
SELECT A PERSONALITY FOR ME
APPLY ENRICHMENT TO:
ALL
CURRENT PAGE
SELECTION
PAGES:
INSERT ENRICHMENT AS:
LINKS
CONTENT
OPTIONS...
808
ZOOM 802
810
PRINT WHAT: DOCUMENT
PAGES PER SHEET: 1 PAGE
PRINT: ALL PAGES IN RANGE
SCALE TO PAPER SIZE: NO SCALING
OPTIONS...
OK
CANCEL
800

FIG. 8

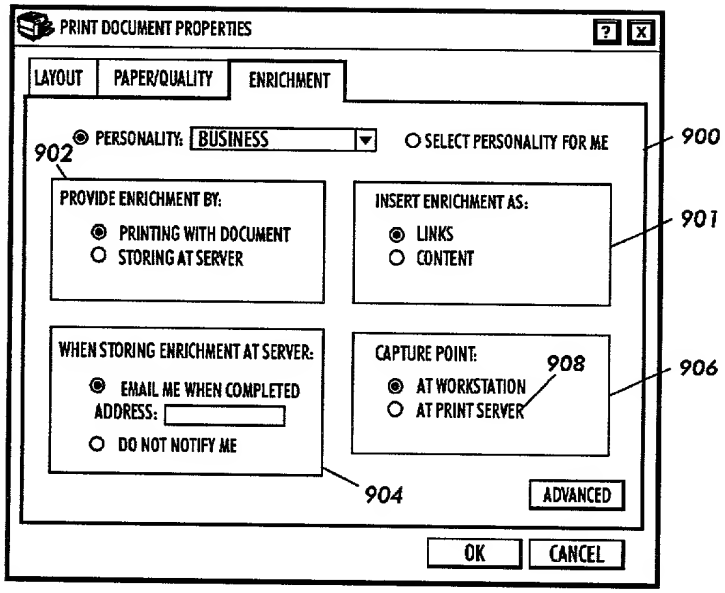
PRINT DOCUMENT PROPERTIES
LAYOUT
PAPER/QUALITY
ENRICHMENT
902
PERSONALITY: BUSINESS
SELECT PERSONALITY FOR ME
900
PROVIDE ENRICHMENT BY:
PRINTING WITH DOCUMENT
STORING AT SERVER
INSERT ENRICHMENT AS:
LINKS
CONTENT
901
WHEN STORING ENRICHMENT AT SERVER:
EMAIL ME WHEN COMPLETED
ADDRESS:
DO NOT NOTIFY ME
904
CAPTURE POINT:
908
AT WORKSTATION
AT PRINT SERVER
906
ADVANCED
OK
CANCEL

FIG. 9

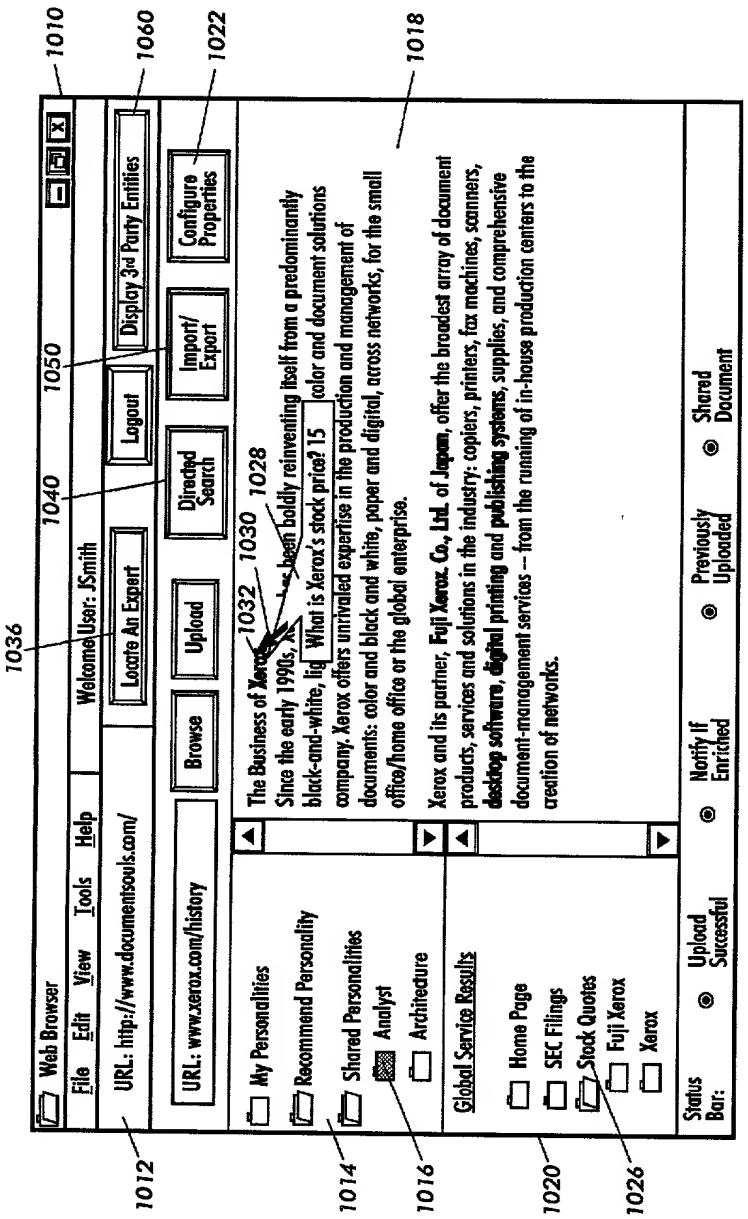
Web Browser
File Edit View Tools Help
Welcome User: JSmith
URL: http://www.documentsouls.com/
Locate An Expert
Logout
Display 3rd Party Entities
URL: www.xerox.com/history
Browse
Upload
Directed Search
Import/ Export
Configure Properties
My Personalities
Recommend Personality
Shared Personalities
Analyst
Architecture
The Business of Xerox
Since the early 1990s, Xerox has been boldly reinventing itself from a predominantly black-and-white, lig
What is Xerox's stock price? 15
color and document solutions company. Xerox offers unrivaled expertise in the production and management of documents: color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.
Xerox and its partner, Fuji Xerox Co., Ltd. of Japan, offer the broadest array of document products, services and solutions in the industry: copiers, printers, fax machines, scanners, desktop software, digital printing and publishing systems, supplies, and comprehensive document-management services -- from the running of in-house production centers to the creation of networks.
Global Service Results
Home Page
SEC Filings
Stock Quotes
Fuji Xerox
Xerox
Status Bar:
Upload Successful
Notify If Enriched
Previously Uploaded
Shared Document
1010
1012
1014
1016
1018
1020
1022
1026
1028
1030
1032
1036
1040
1050
1060

FIG. 10

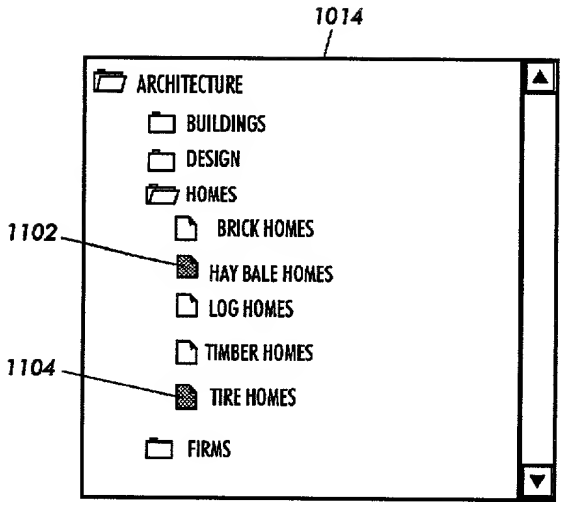
1014
ARCHITECTURE
BUILDINGS
DESIGN
HOMES
BRICK HOMES
1102
HAY BALE HOMES
LOG HOMES
TIMBER HOMES
1104
TIRE HOMES
FIRMS

FIG. 11

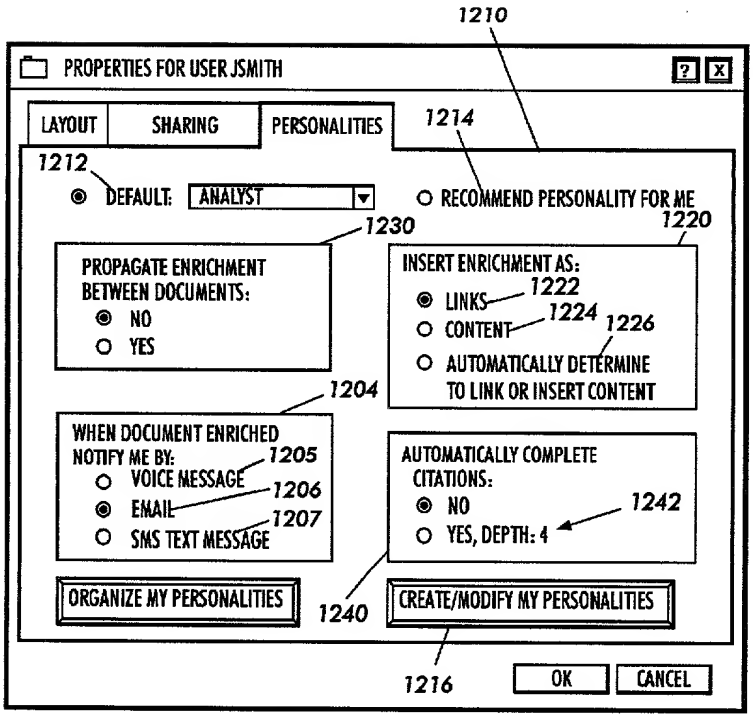
1210
PROPERTIES FOR USER JSMITH
LAYOUT
SHARING
PERSONALITIES
1214
1212
DEFAULT:
ANALYST
RECOMMEND PERSONALITY FOR ME
1230
1220
PROPAGATE ENRICHMENT BETWEEN DOCUMENTS:
NO
YES
INSERT ENRICHMENT AS:
LINKS
1222
CONTENT
1224
1226
AUTOMATICALLY DETERMINE TO LINK OR INSERT CONTENT
1204
WHEN DOCUMENT ENRICHED NOTIFY ME BY:
1205
VOICE MESSAGE
1206
EMAIL
1207
SMS TEXT MESSAGE
AUTOMATICALLY COMPLETE CITATIONS:
NO
1242
YES, DEPTH: 4
ORGANIZE MY PERSONALITIES
1240
CREATE/MODIFY MY PERSONALITIES
1216
OK
CANCEL

FIG. 12

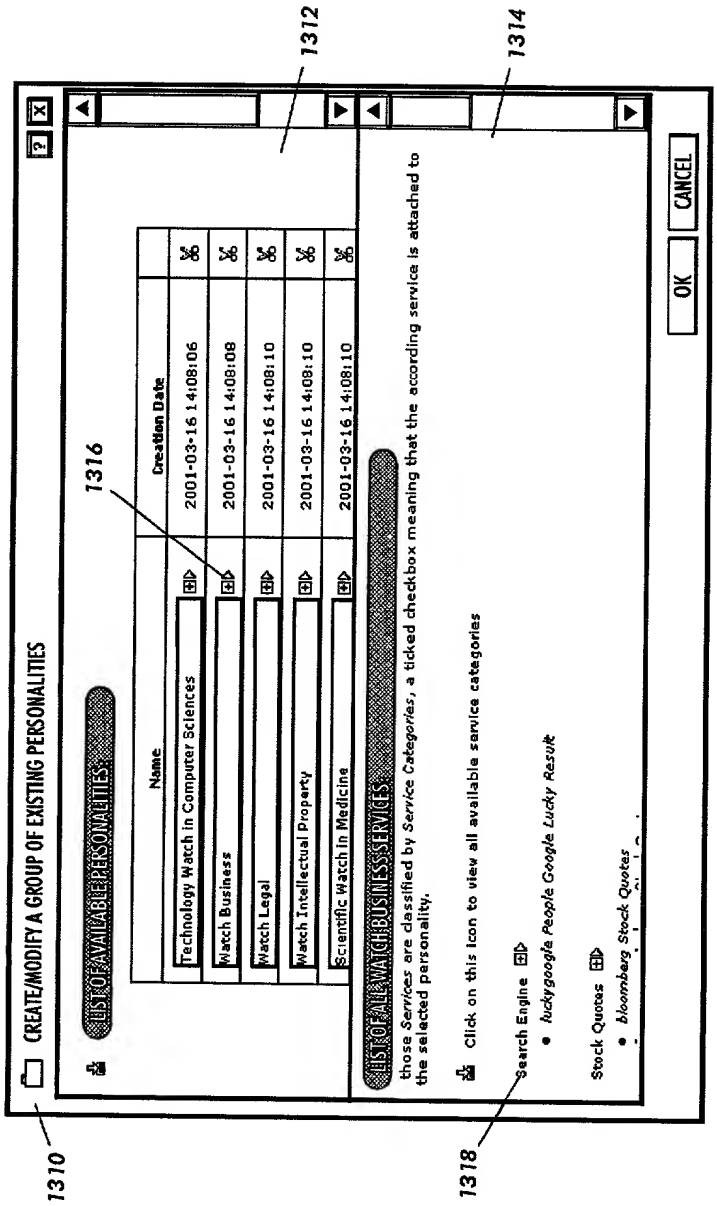
CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES
LIST OF AVAILABLE PERSONALITIES
Name
Creation Date
Technology Watch in Computer Sciences
2001-03-16 14:08:06
Match Business
2001-03-16 14:08:08
Watch Legal
2001-03-16 14:08:10
Watch Intellectual Property
2001-03-16 14:08:10
Scientific Watch in Medicine
2001-03-16 14:08:10
LIST OF ALL WATCH BUSINESS SERVICES
those Services are classified by Service Categories, a ticked checkbox meaning that the according service is attached to the selected personality.
Click on this icon to view all available service categories
Search Engine
luckygoogle People Google Lucky Result
Stock Quotes
Bloomberg Stock Quotes
OK
CANCEL
1310
1312
1314
1316
1318

FIG. 13

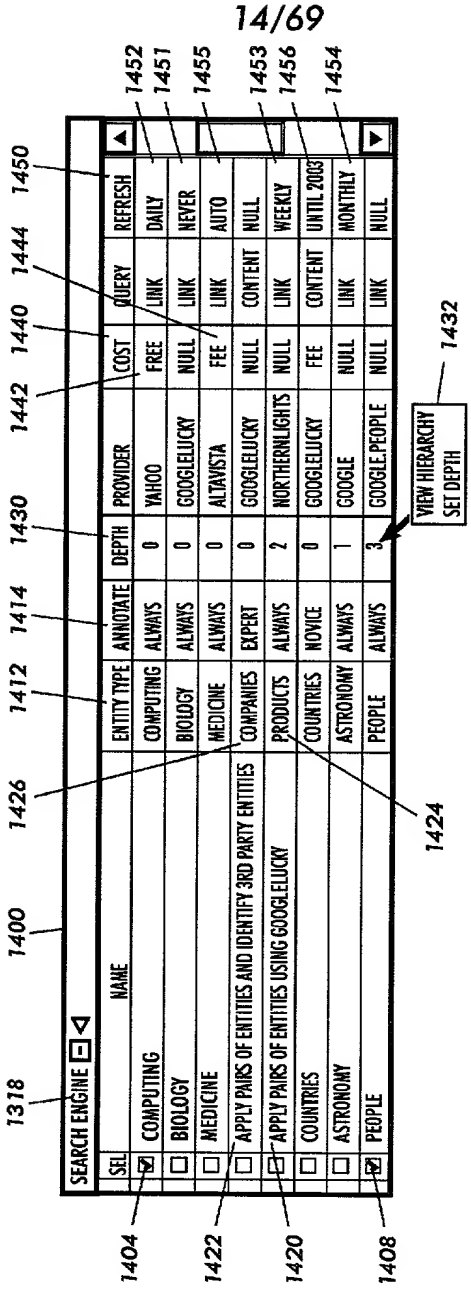
SEARCH ENGINE
SEL
NAME
ENTITY TYPE
ANNOTATE
DEPTH
PROVIDER
COST
QUERY
REFRESH
COMPUTING
COMPUTING
ALWAYS
0
YAHOO
FREE
LINK
DAILY
BIOLOGY
BIOLOGY
ALWAYS
0
GOOGLELUCKY
NULL
LINK
NEVER
MEDICINE
MEDICINE
ALWAYS
0
ALTAVISTA
FEE
LINK
AUTO
APPLY PAIRS OF ENTITIES AND IDENTIFY 3RD PARTY ENTITIES
COMPANIES
EXPERT
0
GOOGLELUCKY
NULL
CONTENT
NULL
APPLY PAIRS OF ENTITIES USING GOOGLELUCKY
PRODUCTS
ALWAYS
2
NORTHERNLIGHTS
NULL
LINK
WEEKLY
COUNTRIES
COUNTRIES
NOVICE
0
GOOGLELUCKY
FEE
CONTENT
UNTIL 2003
ASTRONOMY
ASTRONOMY
ALWAYS
1
GOOGLE
NULL
LINK
MONTHLY
PEOPLE
PEOPLE
ALWAYS
3
GOOGLE.PEOPLE
NULL
LINK
NULL
VIEW HIERARCHY
SET DEPTH
1318
1400
1426
1412
1414
1430
1442
1440
1444
1450
1452
1451
1455
1453
1456
1454
1432
1424
1404
1422
1420
1408

FIG. 14

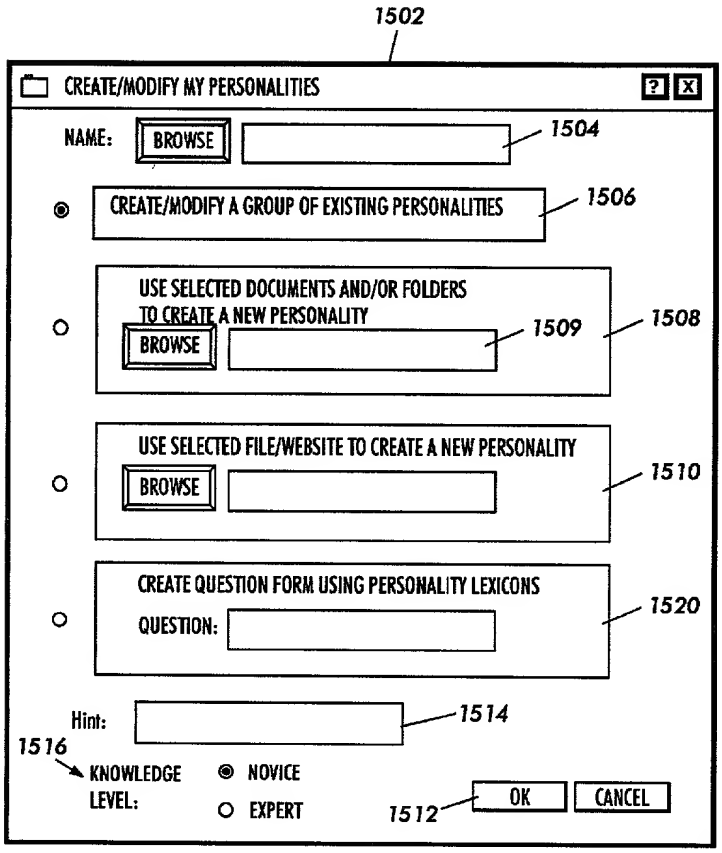
1502
CREATE/MODIFY MY PERSONALITIES
NAME:
BROWSE
1504
CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES
1506
USE SELECTED DOCUMENTS AND/OR FOLDERS TO CREATE A NEW PERSONALITY
BROWSE
1509
1508
USE SELECTED FILE/WEBSITE TO CREATE A NEW PERSONALITY
BROWSE
1510
CREATE QUESTION FORM USING PERSONALITY LEXICONS
QUESTION:
1520
Hint:
1514
1516
KNOWLEDGE LEVEL:
NOVICE
EXPERT
1512
OK
CANCEL

FIG. 15

16/69

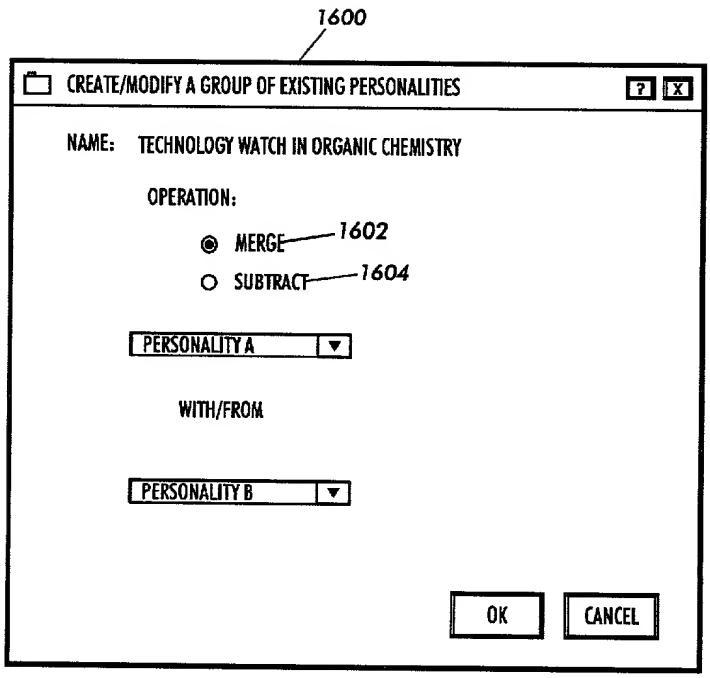
1600
CREATE/MODIFY A GROUP OF EXISTING PERSONALITIES
NAME: TECHNOLOGY WATCH IN ORGANIC CHEMISTRY
OPERATION:
MERGE
1602
SUBTRACT
1604
PERSONALITY A
WITH/FROM
PERSONALITY B
OK
CANCEL

FIG. 16

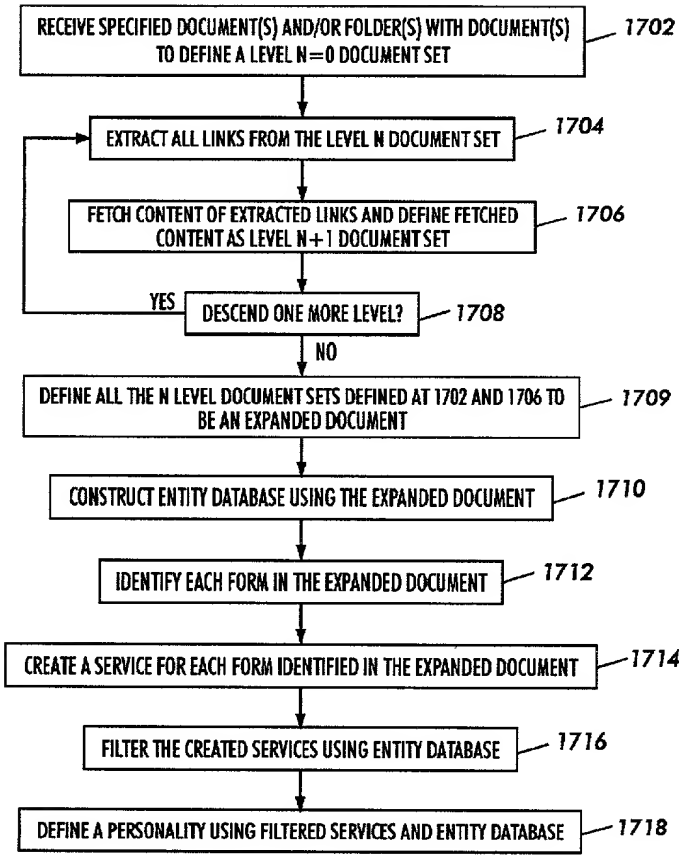
RECEIVE SPECIFIED DOCUMENT(S) AND/OR FOLDER(S) WITH DOCUMENT(S) TO DEFINE A LEVEL N=0 DOCUMENT SET
1702
EXTRACT ALL LINKS FROM THE LEVEL N DOCUMENT SET
1704
FETCH CONTENT OF EXTRACTED LINKS AND DEFINE FETCHED CONTENT AS LEVEL N+1 DOCUMENT SET
1706
YES
DESCEND ONE MORE LEVEL?
1708
NO
DEFINE ALL THE N LEVEL DOCUMENT SETS DEFINED AT 1702 AND 1706 TO BE AN EXPANDED DOCUMENT
1709
CONSTRUCT ENTITY DATABASE USING THE EXPANDED DOCUMENT
1710
IDENTIFY EACH FORM IN THE EXPANDED DOCUMENT
1712
CREATE A SERVICE FOR EACH FORM IDENTIFIED IN THE EXPANDED DOCUMENT
1714
FILTER THE CREATED SERVICES USING ENTITY DATABASE
1716
DEFINE A PERSONALITY USING FILTERED SERVICES AND ENTITY DATABASE
1718

FIG. 17

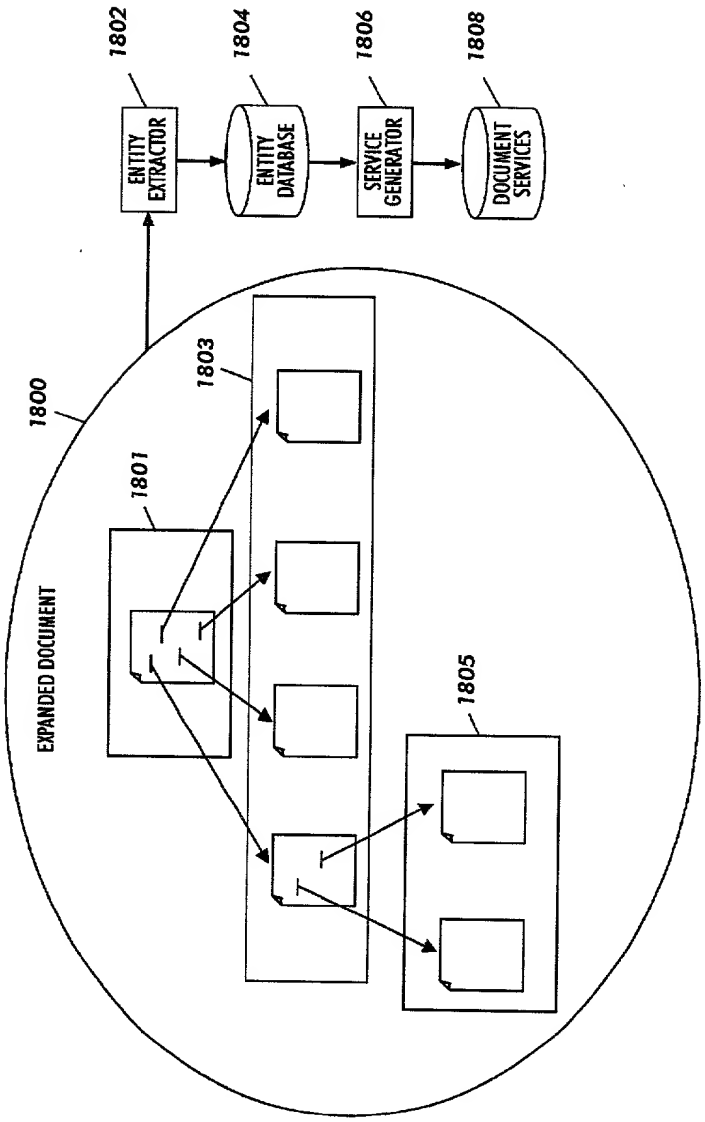
1802
1804
1806
1808
ENTITY EXTRACTOR
ENTITY DATABASE
SERVICE GENERATOR
DOCUMENT SERVICES
1800
1803
1801
EXPANDED DOCUMENT
1805

FIG. 18

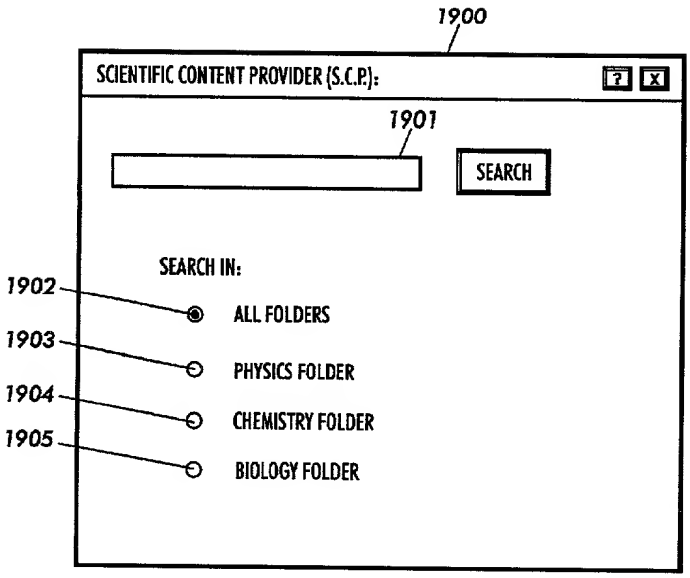
1900
SCIENTIFIC CONTENT PROVIDER (S.C.P.):
1901
SEARCH
SEARCH IN:
1902
ALL FOLDERS
1903
PHYSICS FOLDER
1904
CHEMISTRY FOLDER
1905
BIOLOGY FOLDER

FIG. 19

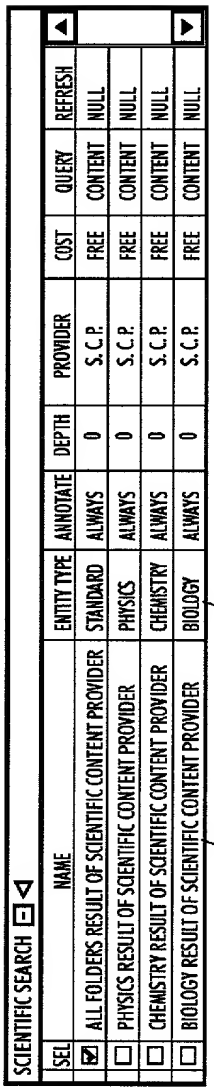
SCIENTIFIC SEARCH
SEL
NAME
ENTITY TYPE
ANNOTATE
DEPTH
PROVIDER
COST
QUERY
REFRESH
ALL FOLDERS RESULT OF SCIENTIFIC CONTENT PROVIDER
STANDARD
ALWAYS
0
S.C.P.
FREE
CONTENT
NULL
PHYSICS RESULT OF SCIENTIFIC CONTENT PROVIDER
PHYSICS
ALWAYS
0
S.C.P.
FREE
CONTENT
NULL
CHEMISTRY RESULT OF SCIENTIFIC CONTENT PROVIDER
CHEMISTRY
ALWAYS
0
S.C.P.
FREE
CONTENT
NULL
BIOLOGY RESULT OF SCIENTIFIC CONTENT PROVIDER
BIOLOGY
ALWAYS
0
S.C.P.
FREE
CONTENT
NULL
2002
2004

FIG. 20

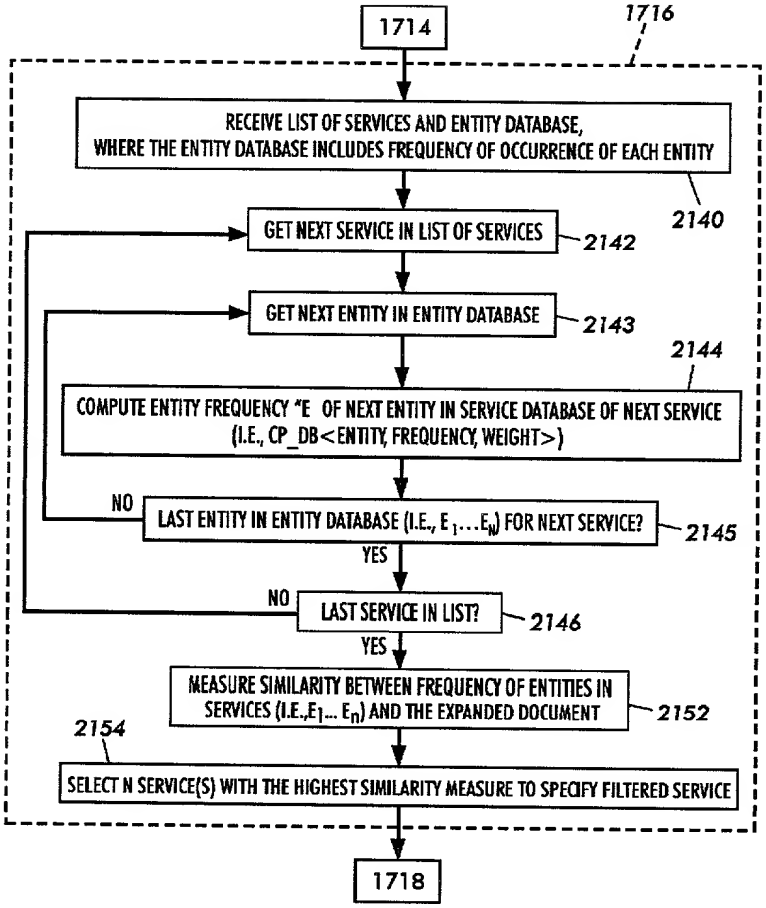
1714
1716
RECEIVE LIST OF SERVICES AND ENTITY DATABASE, WHERE THE ENTITY DATABASE INCLUDES FREQUENCY OF OCCURRENCE OF EACH ENTITY
2140
GET NEXT SERVICE IN LIST OF SERVICES
2142
GET NEXT ENTITY IN ENTITY DATABASE
2143
2144
COMPUTE ENTITY FREQUENCY "E OF NEXT ENTITY IN SERVICE DATABASE OF NEXT SERVICE (I.E., CP_DB<ENTITY, FREQUENCY, WEIGHT>)
NO
LAST ENTITY IN ENTITY DATABASE (I.E., E1...EN) FOR NEXT SERVICE?
2145
YES
NO
LAST SERVICE IN LIST?
2146
YES
MEASURE SIMILARITY BETWEEN FREQUENCY OF ENTITIES IN SERVICES (I.E.,E1... En) AND THE EXPANDED DOCUMENT
2152
2154
SELECT N SERVICE(S) WITH THE HIGHEST SIMILARITY MEASURE TO SPECIFY FILTERED SERVICE
1718

FIG. 21

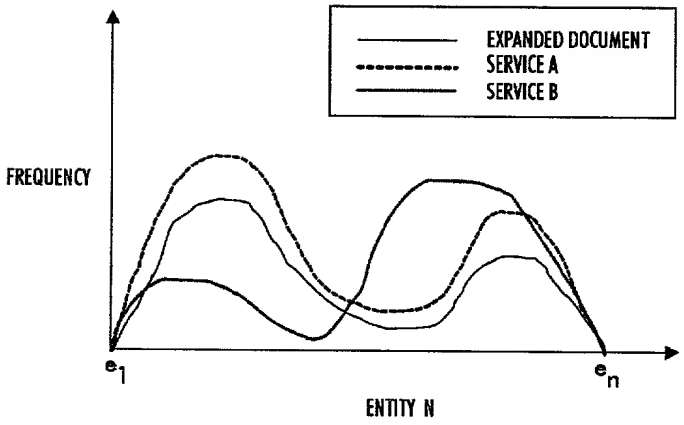
EXPANDED DOCUMENT
SERVICE A
SERVICE B
FREQUENCY
e_1
e_n
ENTITY N

FIG. 22

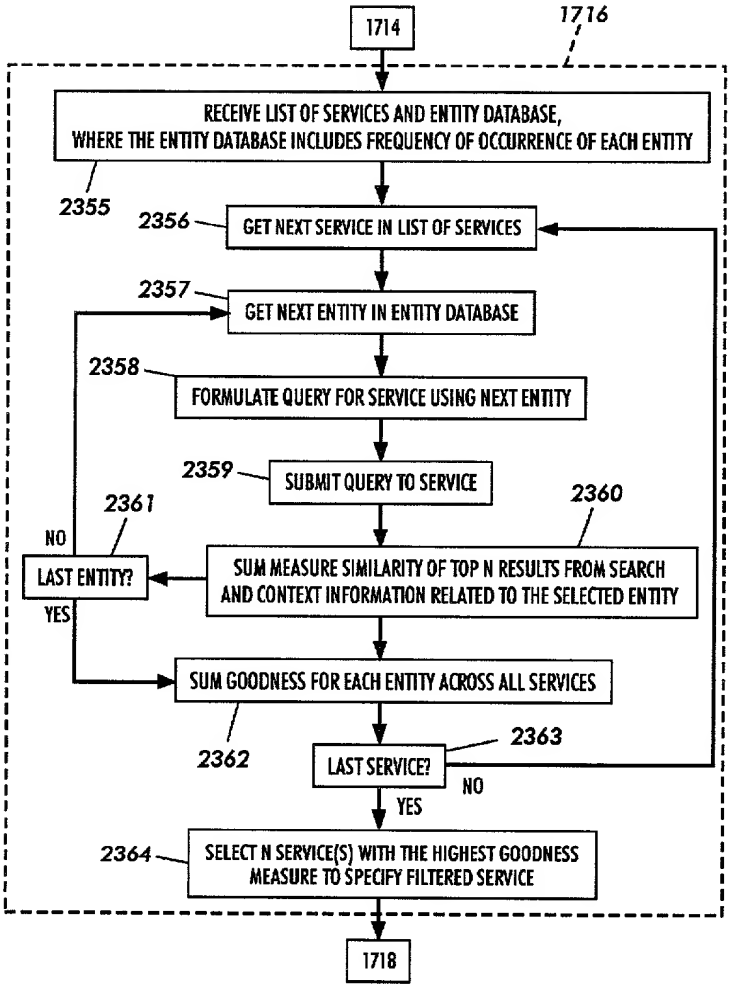
1714
1716
RECEIVE LIST OF SERVICES AND ENTITY DATABASE, WHERE THE ENTITY DATABASE INCLUDES FREQUENCY OF OCCURRENCE OF EACH ENTITY
2355
2356
GET NEXT SERVICE IN LIST OF SERVICES
2357
GET NEXT ENTITY IN ENTITY DATABASE
2358
FORMULATE QUERY FOR SERVICE USING NEXT ENTITY
2359
SUBMIT QUERY TO SERVICE
2360
2361
NO
LAST ENTITY?
SUM MEASURE SIMILARITY OF TOP N RESULTS FROM SEARCH AND CONTEXT INFORMATION RELATED TO THE SELECTED ENTITY
YES
SUM GOODNESS FOR EACH ENTITY ACROSS ALL SERVICES
2362
2363
LAST SERVICE?
NO
YES
2364
SELECT N SERVICE(S) WITH THE HIGHEST GOODNESS MEASURE TO SPECIFY FILTERED SERVICE
1718

FIG. 23

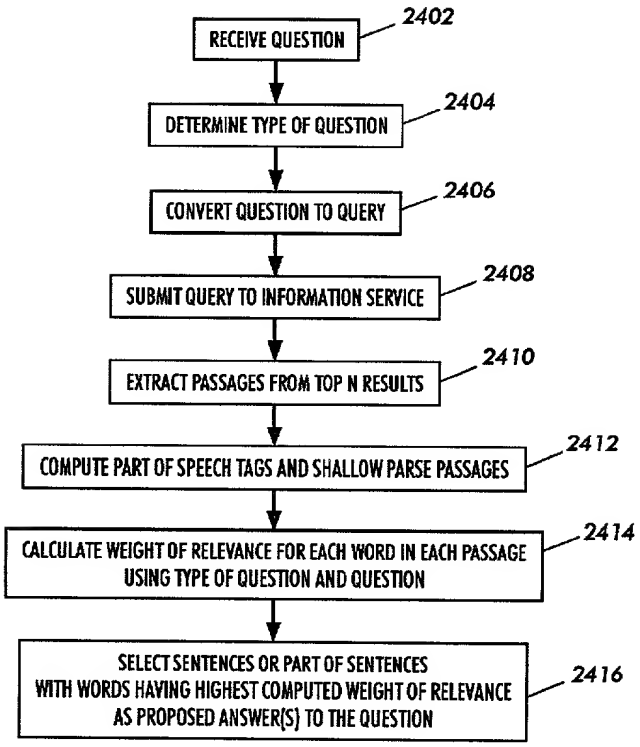
RECEIVE QUESTION
2402
DETERMINE TYPE OF QUESTION
2404
CONVERT QUESTION TO QUERY
2406
SUBMIT QUERY TO INFORMATION SERVICE
2408
EXTRACT PASSAGES FROM TOP N RESULTS
2410
COMPUTE PART OF SPEECH TAGS AND SHALLOW PARSE PASSAGES
2412
CALCULATE WEIGHT OF RELEVANCE FOR EACH WORD IN EACH PASSAGE USING TYPE OF QUESTION AND QUESTION
2414
SELECT SENTENCES OR PART OF SENTENCES WITH WORDS HAVING HIGHEST COMPUTED WEIGHT OF RELEVANCE AS PROPOSED ANSWER(S) TO THE QUESTION
2416

FIG. 24

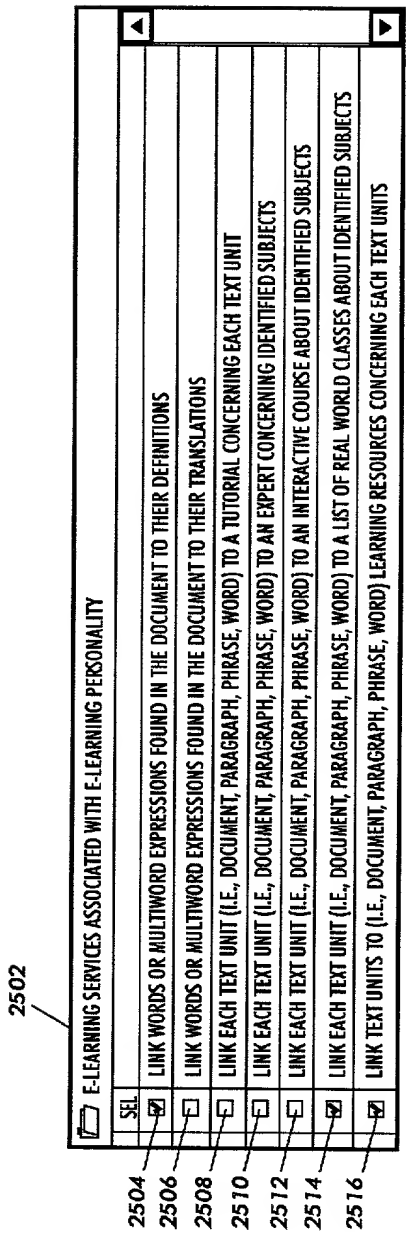
2502
E-LEARNING SERVICES ASSOCIATED WITH E-LEARNING PERSONALITY
SEL
LINK WORDS OR MULTIWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR DEFINITIONS
LINK WORDS OR MULTIWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR TRANSLATIONS
LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A TUTORIAL CONCERNING EACH TEXT UNIT
LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO AN EXPERT CONCERNING IDENTIFIED SUBJECTS
LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO AN INTERACTIVE COURSE ABOUT IDENTIFIED SUBJECTS
LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A LIST OF REAL WORLD CLASSES ABOUT IDENTIFIED SUBJECTS
LINK TEXT UNITS TO (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) LEARNING RESOURCES CONCERNING EACH TEXT UNITS
2504
2506
2508
2510
2512
2514
2516

FIG. 25

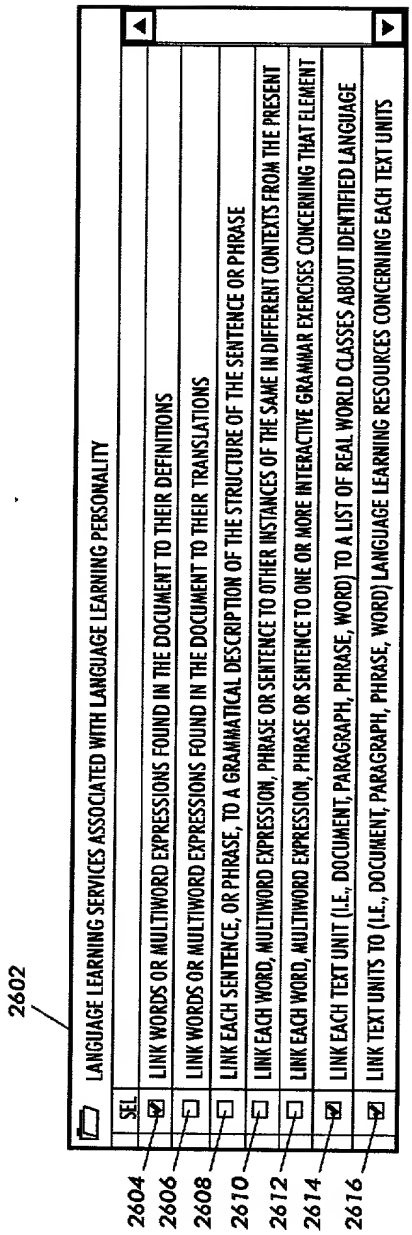
2602
LANGUAGE LEARNING SERVICES ASSOCIATED WITH LANGUAGE LEARNING PERSONALITY
SEL
LINK WORDS OR MULTIWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR DEFINITIONS
LINK WORDS OR MULTIWORD EXPRESSIONS FOUND IN THE DOCUMENT TO THEIR TRANSLATIONS
LINK EACH SENTENCE, OR PHRASE, TO A GRAMMATICAL DESCRIPTION OF THE STRUCTURE OF THE SENTENCE OR PHRASE
LINK EACH WORD, MULTIWORD EXPRESSION, PHRASE OR SENTENCE TO OTHER INSTANCES OF THE SAME IN DIFFERENT CONTEXTS FROM THE PRESENT
LINK EACH WORD, MULTIWORD EXPRESSION, PHRASE OR SENTENCE TO ONE OR MORE INTERACTIVE GRAMMAR EXERCISES CONCERNING THAT ELEMENT
LINK EACH TEXT UNIT (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) TO A LIST OF REAL WORLD CLASSES ABOUT IDENTIFIED LANGUAGE
LINK TEXT UNITS TO (I.E., DOCUMENT, PARAGRAPH, PHRASE, WORD) LANGUAGE LEARNING RESOURCES CONCERNING EACH TEXT UNITS
2604
2606
2608
2610
2612
2614
2616

FIG. 26

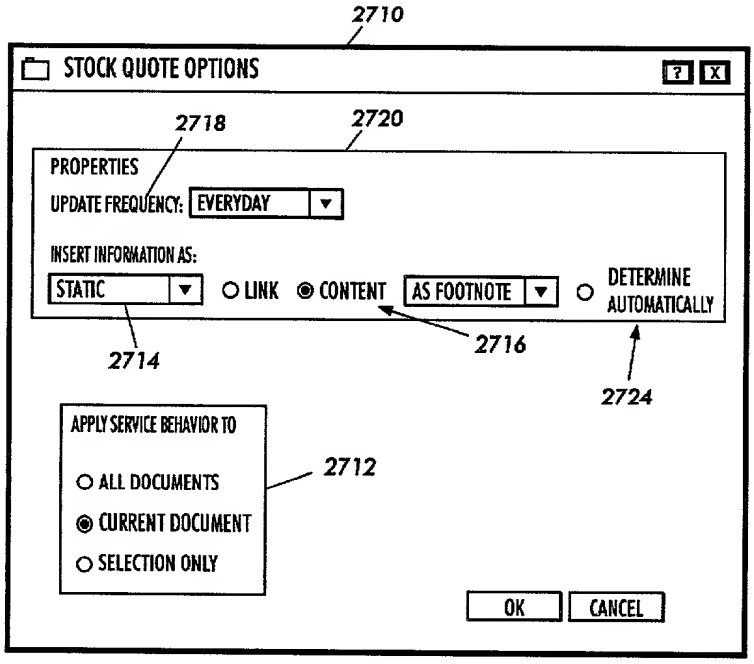
2710
STOCK QUOTE OPTIONS
2718
2720
PROPERTIES
UPDATE FREQUENCY: EVERYDAY
INSERT INFORMATION AS:
STATIC
LINK
CONTENT
AS FOOTNOTE
DETERMINE AUTOMATICALLY
2714
2716
2724
APPLY SERVICE BEHAVIOR TO
ALL DOCUMENTS
CURRENT DOCUMENT
SELECTION ONLY
2712
OK
CANCEL

FIG. 27

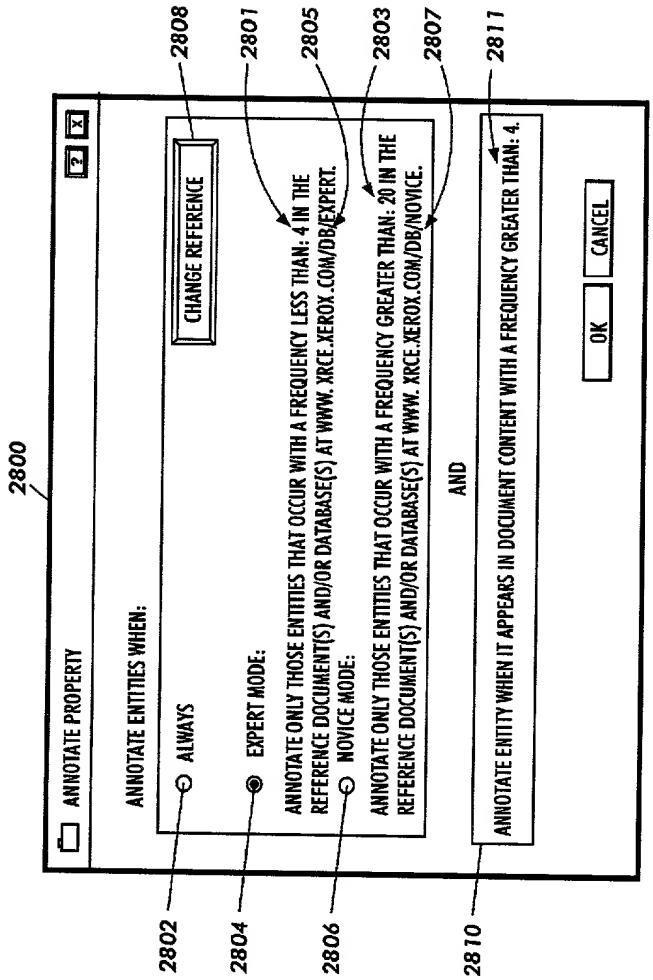
2800
ANNOTATE PROPERTY
ANNOTATE ENTITIES WHEN:
ALWAYS
EXPERT MODE:
NOVICE MODE:
CHANGE REFERENCE
ANNOTATE ONLY THOSE ENTITIES THAT OCCUR WITH A FREQUENCY LESS THAN: 4 IN THE REFERENCE DOCUMENT(S) AND/OR DATABASE(S) AT WWW.XRCE.XEROX.COM/DB/EXPERT.
ANNOTATE ONLY THOSE ENTITIES THAT OCCUR WITH A FREQUENCY GREATER THAN: 20 IN THE REFERENCE DOCUMENT(S) AND/OR DATABASE(S) AT WWW.XRCE.XEROX.COM/DB/NOVICE.
AND
ANNOTATE ENTITY WHEN IT APPEARS IN DOCUMENT CONTENT WITH A FREQUENCY GREATER THAN: 4.
OK
CANCEL
2802
2804
2806
2808
2801
2805
2803
2807
2810
2811

FIG. 28

29/69

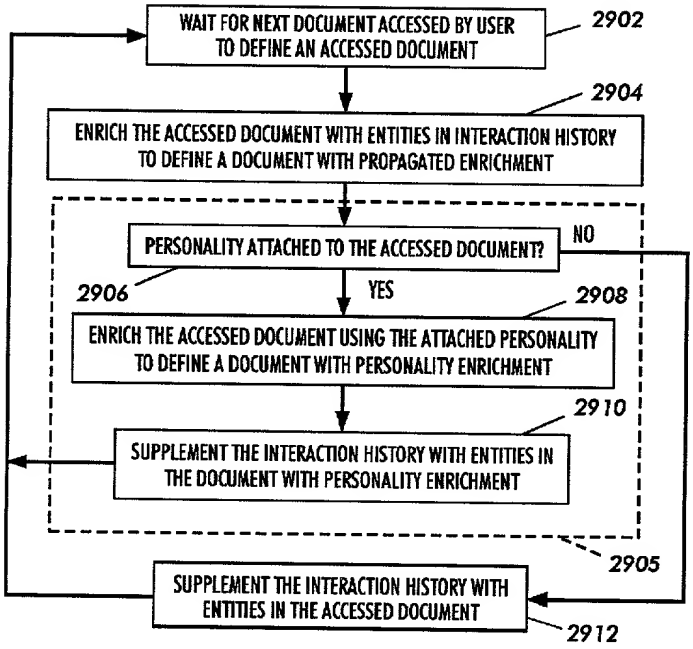
WAIT FOR NEXT DOCUMENT ACCESSED BY USER TO DEFINE AN ACCESSED DOCUMENT
2902
ENRICH THE ACCESSED DOCUMENT WITH ENTITIES IN INTERACTION HISTORY TO DEFINE A DOCUMENT WITH PROPAGATED ENRICHMENT
2904
PERSONALITY ATTACHED TO THE ACCESSED DOCUMENT?
NO
YES
2906
ENRICH THE ACCESSED DOCUMENT USING THE ATTACHED PERSONALITY TO DEFINE A DOCUMENT WITH PERSONALITY ENRICHMENT
2908
SUPPLEMENT THE INTERACTION HISTORY WITH ENTITIES IN THE DOCUMENT WITH PERSONALITY ENRICHMENT
2910
2905
SUPPLEMENT THE INTERACTION HISTORY WITH ENTITIES IN THE ACCESSED DOCUMENT
2912

FIG. 29

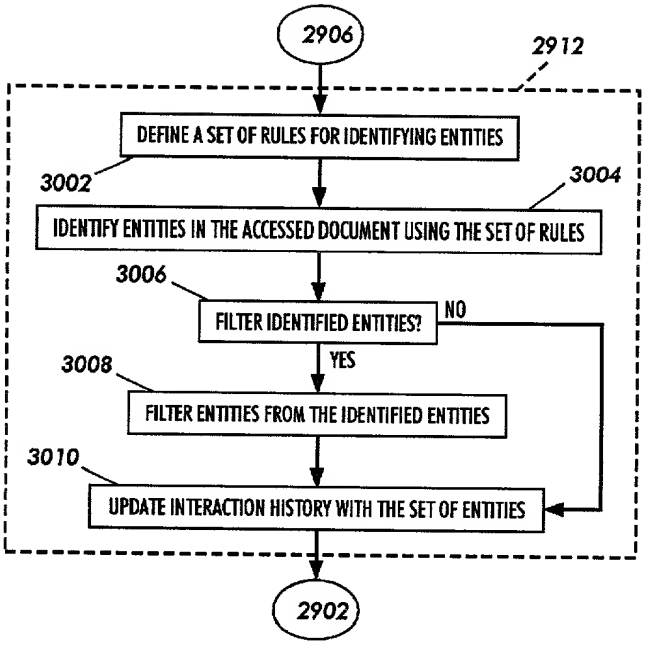
2906
2912
DEFINE A SET OF RULES FOR IDENTIFYING ENTITIES
3002
3004
IDENTIFY ENTITIES IN THE ACCESSED DOCUMENT USING THE SET OF RULES
3006
FILTER IDENTIFIED ENTITIES?
NO
YES
3008
FILTER ENTITIES FROM THE IDENTIFIED ENTITIES
3010
UPDATE INTERACTION HISTORY WITH THE SET OF ENTITIES
2902

FIG. 30

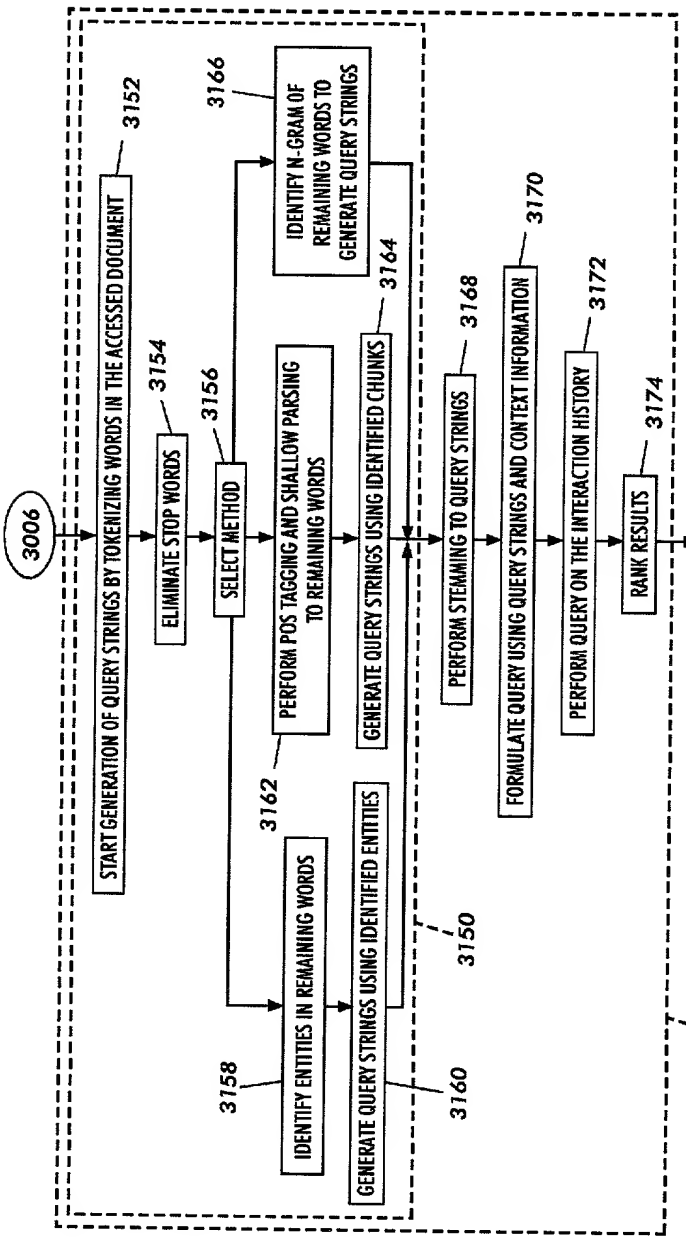
3006
START GENERATION OF QUERY STRINGS BY TOKENIZING WORDS IN THE ACCESSED DOCUMENT
3152
ELIMINATE STOP WORDS
3154
SELECT METHOD
3156
IDENTIFY ENTITIES IN REMAINING WORDS
3158
GENERATE QUERY STRINGS USING IDENTIFIED ENTITIES
3160
PERFORM POS TAGGING AND SHALLOW PARSING TO REMAINING WORDS
3162
IDENTIFY N-GRAM OF REMAINING WORDS TO GENERATE QUERY STRINGS
3166
GENERATE QUERY STRINGS USING IDENTIFIED CHUNKS
3164
PERFORM STEMMING TO QUERY STRINGS
3168
FORMULATE QUERY USING QUERY STRINGS AND CONTEXT INFORMATION
3170
PERFORM QUERY ON THE INTERACTION HISTORY
3172
RANK RESULTS
3174
3010
3150
3008

FIG. 31

32/69

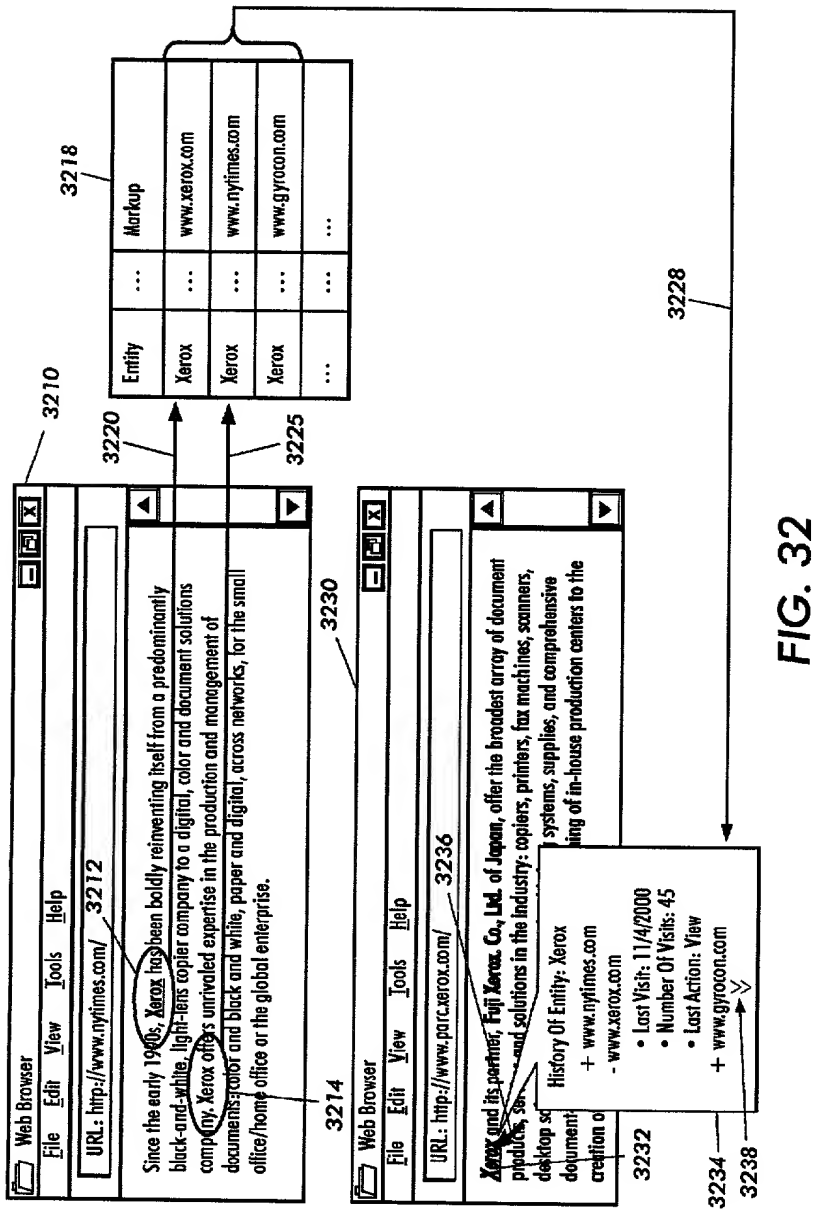
3210
3212
3214
3218
3220
3225
3228
3230
3232
3234
3236
3238
Web Browser
File Edit View Tools Help
URL: http://www.nytimes.com/
Since the early 1960s, Xerox has been boldly reinventing itself from a predominantly black-and-white, light-lens copier company to a digital, color and document solutions company. Xerox offers unrivaled expertise in the production and management of documents—color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.
Entity | ... | Markup
Xerox | ... | www.xerox.com
Xerox | ... | www.nytimes.com
Xerox | ... | www.gyrocon.com
... | ... | ...
URL: http://www.parc.xerox.com/
Xerox and its partner, Fuji Xerox Co., Ltd. of Japan, offer the broadest array of document
History Of Entity: Xerox
+ www.nytimes.com
- www.xerox.com
• Last Visit: 11/4/2000
• Number Of Visits: 45
• Last Action: View
+ www.gyrocon.com

FIG. 32

3218

ENTITY	ENTITY TYPE	VISITS	LAST VISIT	ACTION LIST	DMOZ CATEGORY	LOCATION	POS	MARKUP
XEROX	BUSINESS	45	11/4/2000	VIEW (43), PRINT(2)	BUSINESS	40	NOUN	WWW.XEROX.COM
XEROX	BUSINESS	1	5/1/2001	VIEW(1)	FINANCE	689	NOUN	WWW.NYTIMES.COM
XEROX	BUSINESS	4	2/22/2001	EMAIL(1), VIEW(3)	COMPUTER	343	NOUN	WWW.GYROCON.COM
...	...	...	...	...	...	...	...	...

FIG. 33

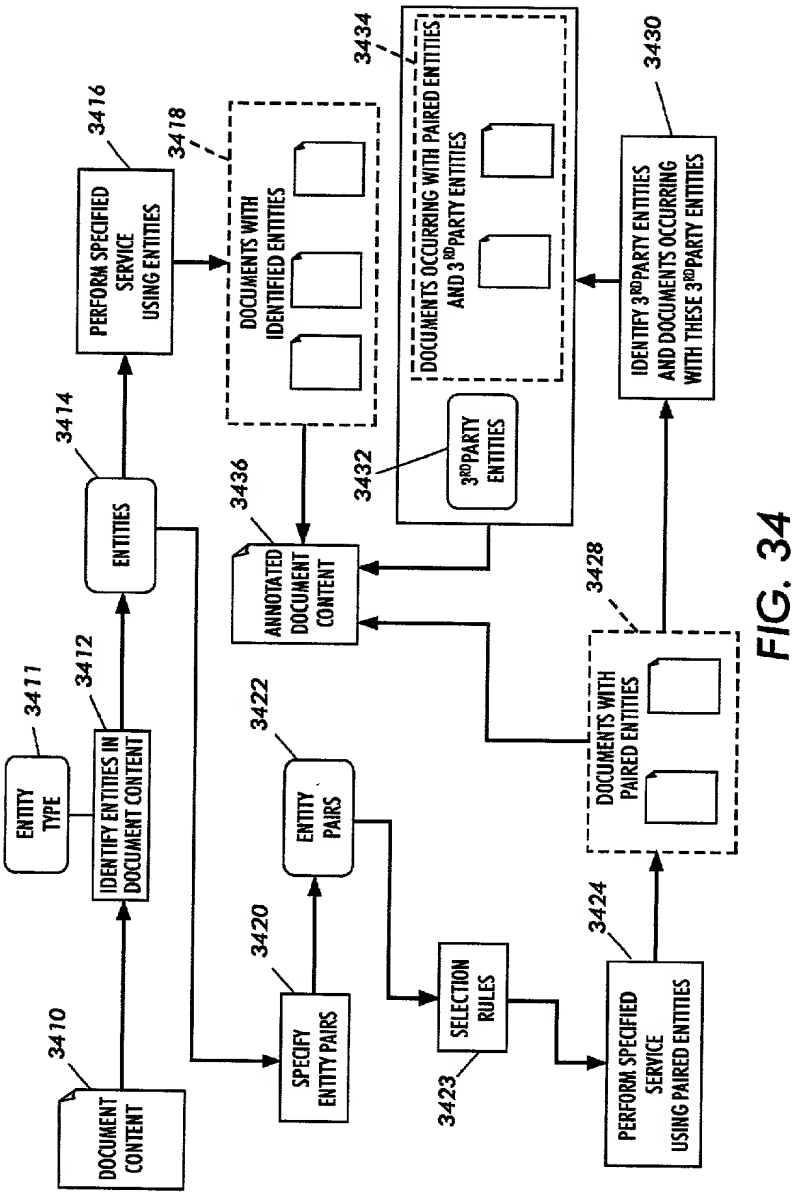
DOCUMENT CONTENT
3410
ENTITY TYPE
3411
IDENTIFY ENTITIES IN DOCUMENT CONTENT
3412
ENTITIES
3414
PERFORM SPECIFIED SERVICE USING ENTITIES
3416
DOCUMENTS WITH IDENTIFIED ENTITIES
3418
SPECIFY ENTITY PAIRS
3420
ENTITY PAIRS
3422
SELECTION RULES
3423
PERFORM SPECIFIED SERVICE USING PAIRED ENTITIES
3424
ANNOTATED DOCUMENT CONTENT
3436
3RD PARTY ENTITIES
3432
DOCUMENTS OCCURRING WITH PAIRED ENTITIES AND 3RD PARTY ENTITIES
3434
DOCUMENTS WITH PAIRED ENTITIES
3428
IDENTIFY 3RD PARTY ENTITIES AND DOCUMENTS OCCURRING WITH THESE 3RD PARTY ENTITIES
3430

FIG. 34

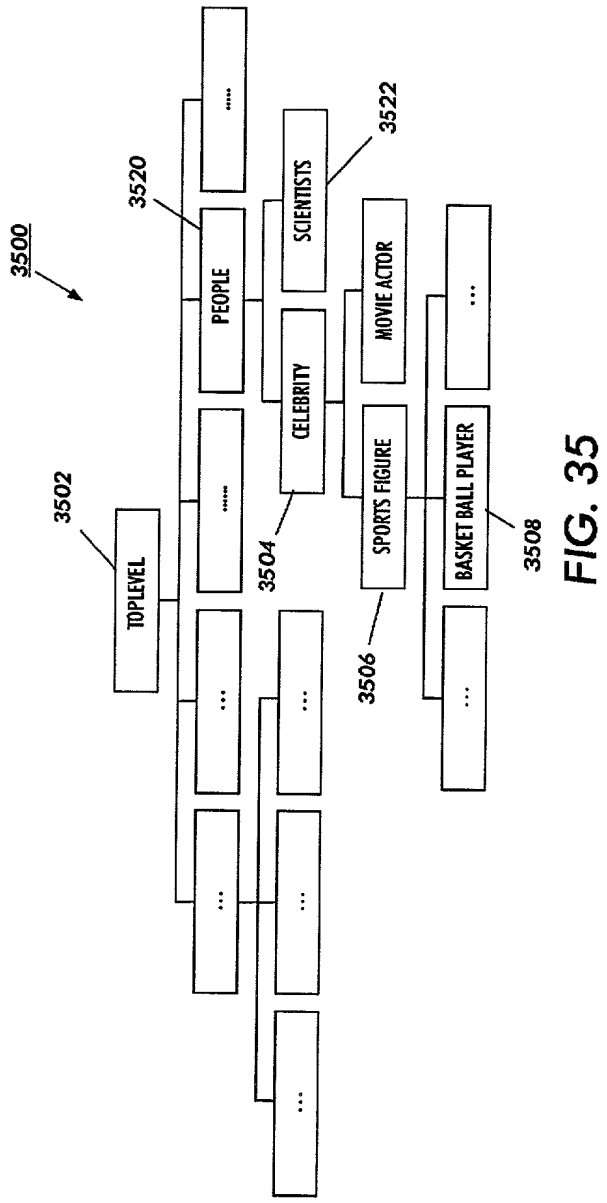
3500
3502
3504
3506
3508
3520
3522
TOPLEVEL
PEOPLE
CELEBRITY
SCIENTISTS
MOVIE ACTOR
SPORTS FIGURE
BASKET BALL PLAYER

FIG. 35

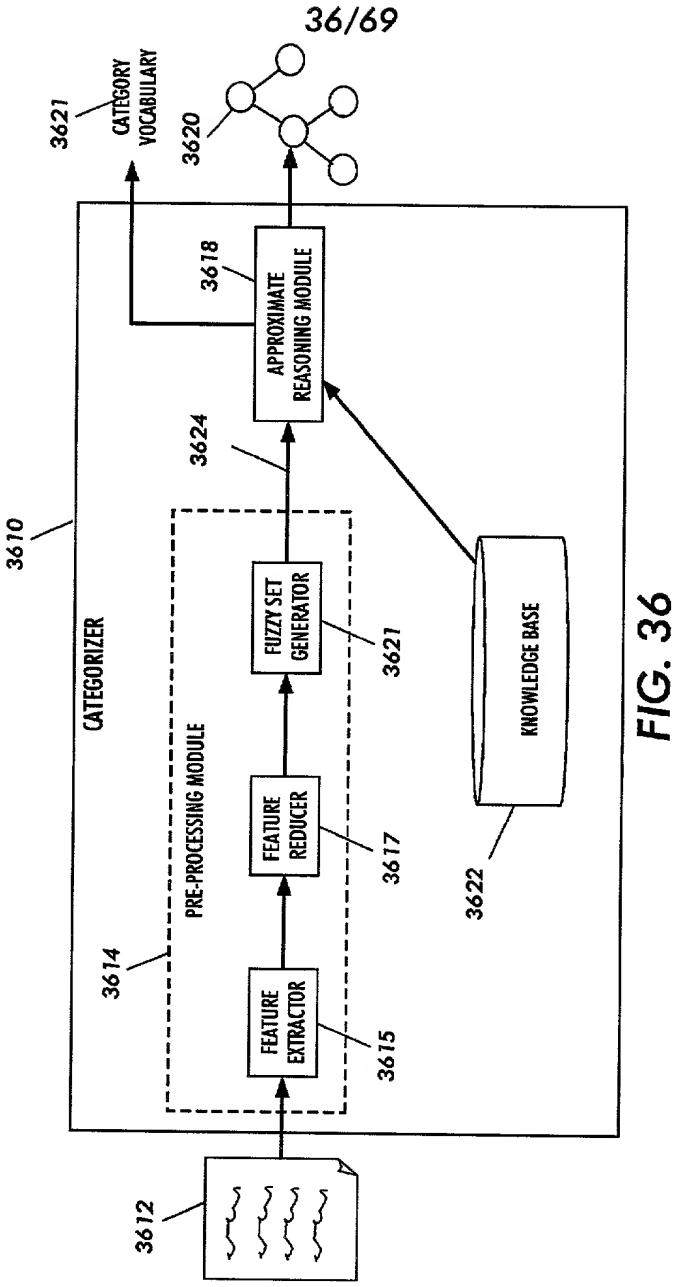
36/69
3621
CATEGORY VOCABULARY
3620
3618
APPROXIMATE REASONING MODULE
3624
3610
CATEGORIZER
PRE-PROCESSING MODULE
FUZZY SET GENERATOR
3621
FEATURE REDUCER
3617
FEATURE EXTRACTOR
3615
3614
KNOWLEDGE BASE
3622
3612

FIG. 36

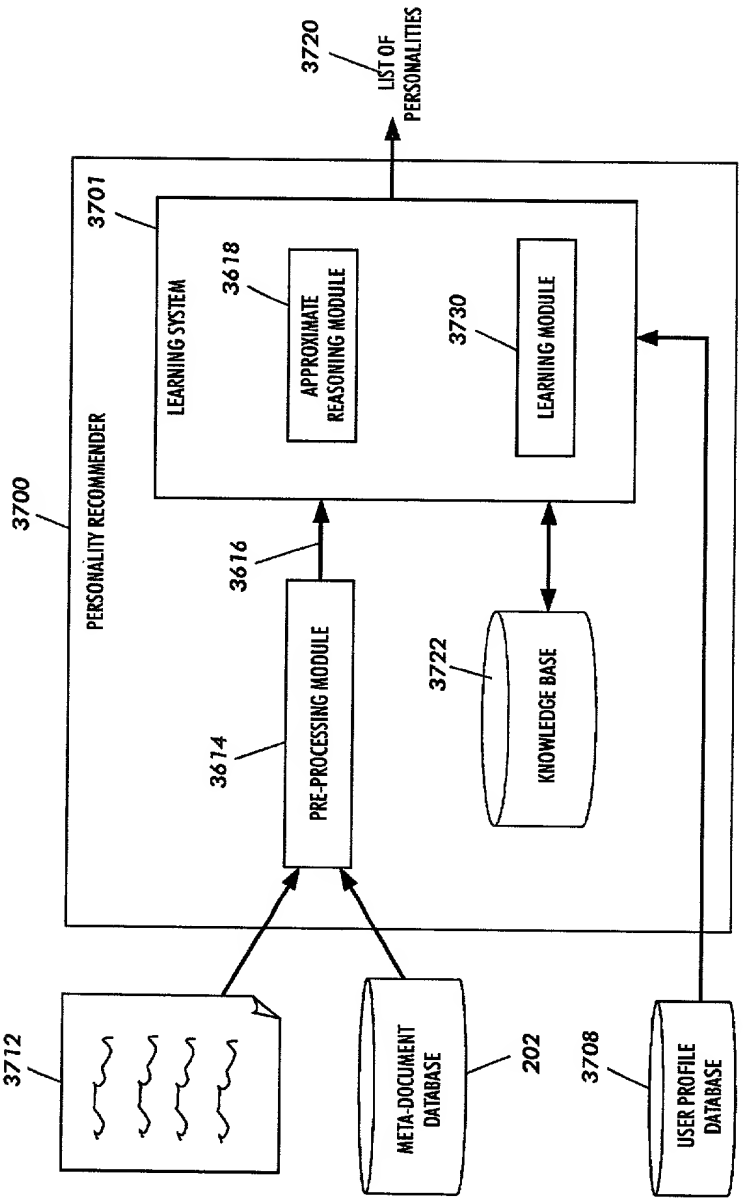
3720
LIST OF PERSONALITIES
3701
LEARNING SYSTEM
3618
APPROXIMATE REASONING MODULE
3730
LEARNING MODULE
3700
PERSONALITY RECOMMENDER
3616
PRE-PROCESSING MODULE
3722
KNOWLEDGE BASE
3614
3712
META-DOCUMENT DATABASE
202
3708
USER PROFILE DATABASE

FIG. 37

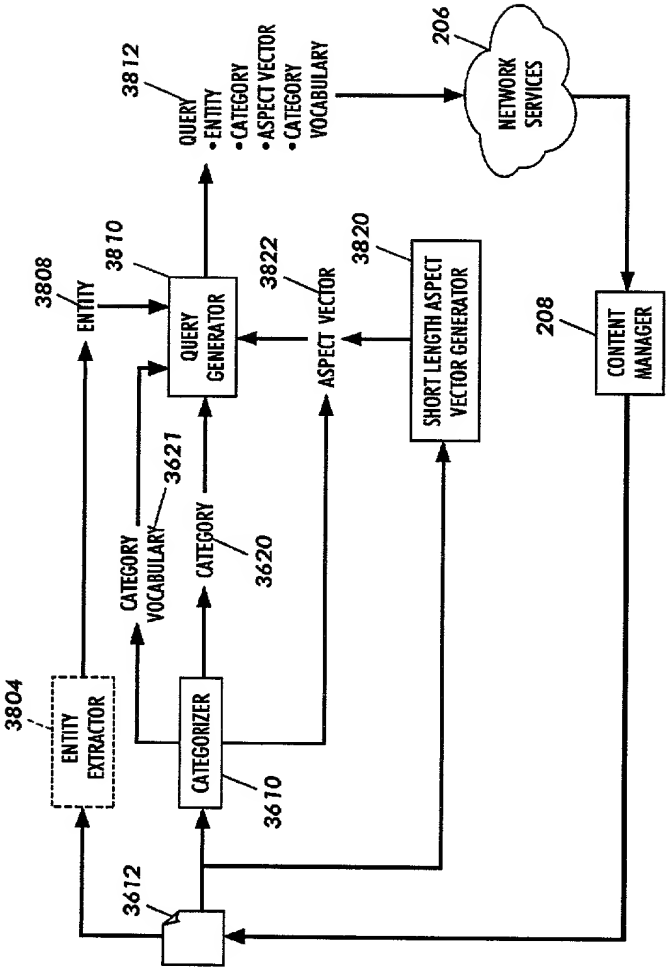
3812
QUERY
•ENTITY
•CATEGORY
•ASPECT VECTOR
•CATEGORY VOCABULARY
206
NETWORK SERVICES
3808
ENTITY
3810
QUERY GENERATOR
3822
ASPECT VECTOR
3820
SHORT LENGTH ASPECT VECTOR GENERATOR
208
CONTENT MANAGER
3621
CATEGORY VOCABULARY
CATEGORY
3620
3804
ENTITY EXTRACTOR
CATEGORIZER
3610
3612

FIG. 38

39/69

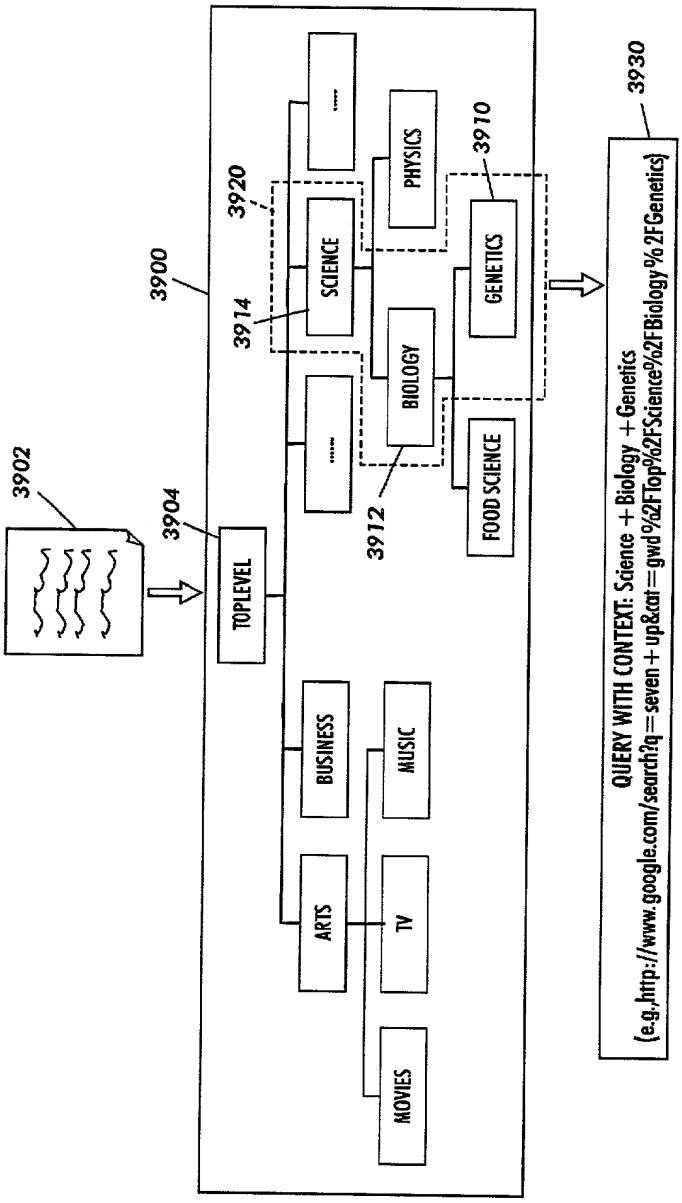
3902
3904
3900
TOPLEVEL
3920
3914
3912
3910
3930
SCIENCE
.....
PHYSICS
BIOLOGY
GENETICS
FOOD SCIENCE
BUSINESS
MUSIC
ARTS
TV
MOVIES
QUERY WITH CONTEXT: Science + Biology + Genetics
(e.g.,http://www.google.com/search?q=seven+up&aq=gwd%2Ftop%2Fscience%2FBiology%2FGenetics)

FIG. 39

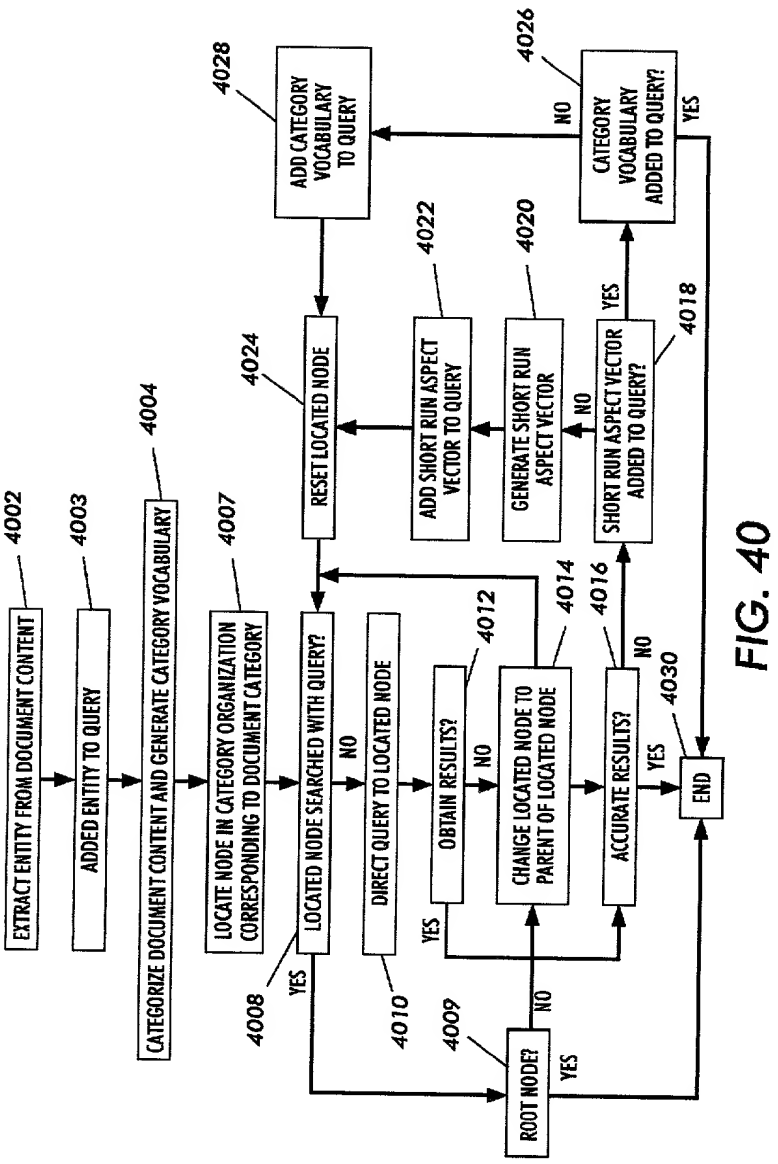
EXTRACT ENTITY FROM DOCUMENT CONTENT
4002
ADDED ENTITY TO QUERY
4003
CATEGORIZE DOCUMENT CONTENT AND GENERATE CATEGORY VOCABULARY
4004
LOCATE NODE IN CATEGORY ORGANIZATION CORRESPONDING TO DOCUMENT CATEGORY
4007
LOCATED NODE SEARCHED WITH QUERY?
4008
YES
NO
DIRECT QUERY TO LOCATED NODE
4010
OBTAIN RESULTS?
4012
YES
NO
CHANGE LOCATED NODE TO PARENT OF LOCATED NODE
4014
ACCURATE RESULTS?
4016
YES
NO
ROOT NODE?
4009
NO
YES
END
4030
RESET LOCATED NODE
4024
ADD SHORT RUN ASPECT VECTOR TO QUERY
4022
GENERATE SHORT RUN ASPECT VECTOR
4020
SHORT RUN ASPECT VECTOR ADDED TO QUERY?
4018
YES
NO
CATEGORY VOCABULARY ADDED TO QUERY?
4026
YES
NO
ADD CATEGORY VOCABULARY TO QUERY
4028

FIG. 40

41/69

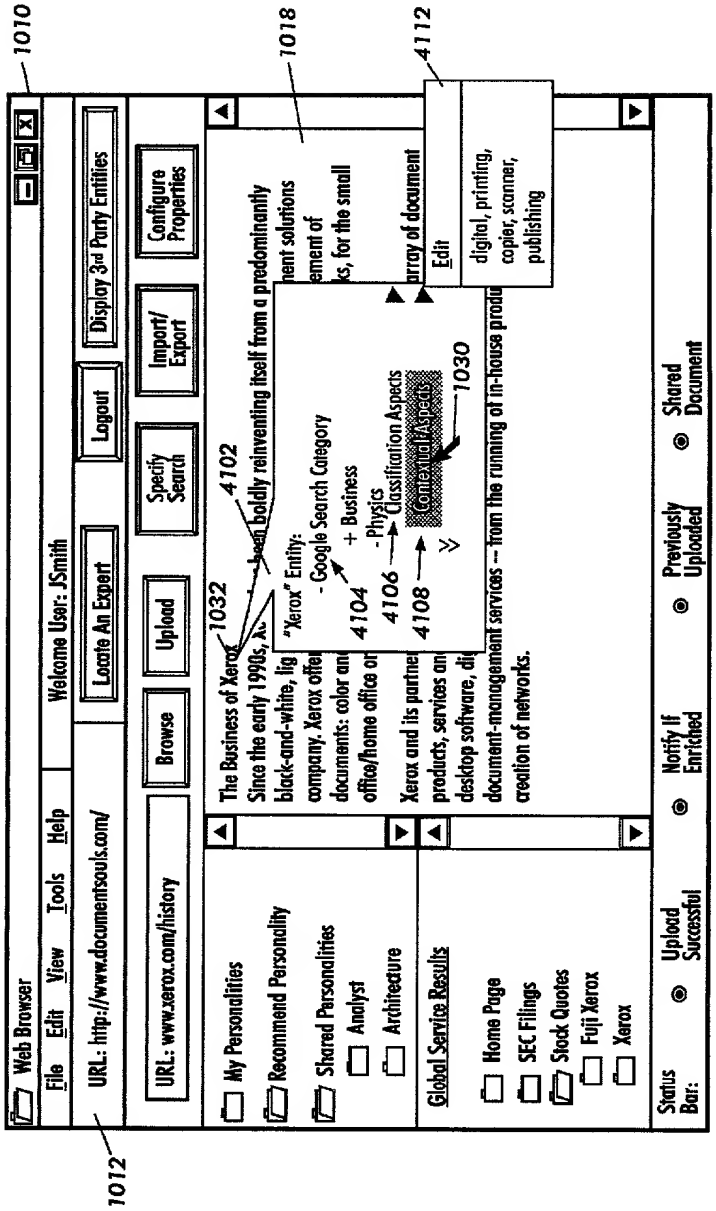
1010
1012
1018
4112
Web Browser
File Edit View Tools Help
Welcome User: JSmith
URL: http://www.documentsouls.com/
Locate An Expert
Logout
Display 3rd Party Entities
URL: www.xerox.com/history
Browse
Upload
Specify Search
Import/ Export
Configure Properties
My Personalities
Recommend Personality
Shared Personalities
Analyst
Architecture
Global Service Results
Home Page
SEC Filings
Stock Quotes
Fuji Xerox
Xerox
The Business of Xerox
Since the early 1990s, Xe
black-and-white, lig
company. Xerox offe
documents: color an
office/home office or
Xerox and its partner
products, services an
desktop software, di
document-management services -- from the running of in-house produ
creation of networks.
1032
4102
"Xerox" Entity:
- Google Search Category
+ Business
- Physics
4104
4106
Classification Aspects
Content Aspects
4108
1030
ement of
ks, for the small
array of document
Edit
digital, printing,
copier, scanner,
publishing
Status Bar:
Upload Successful
Notify If Enriched
Previously Uploaded
Shared Document

FIG. 41

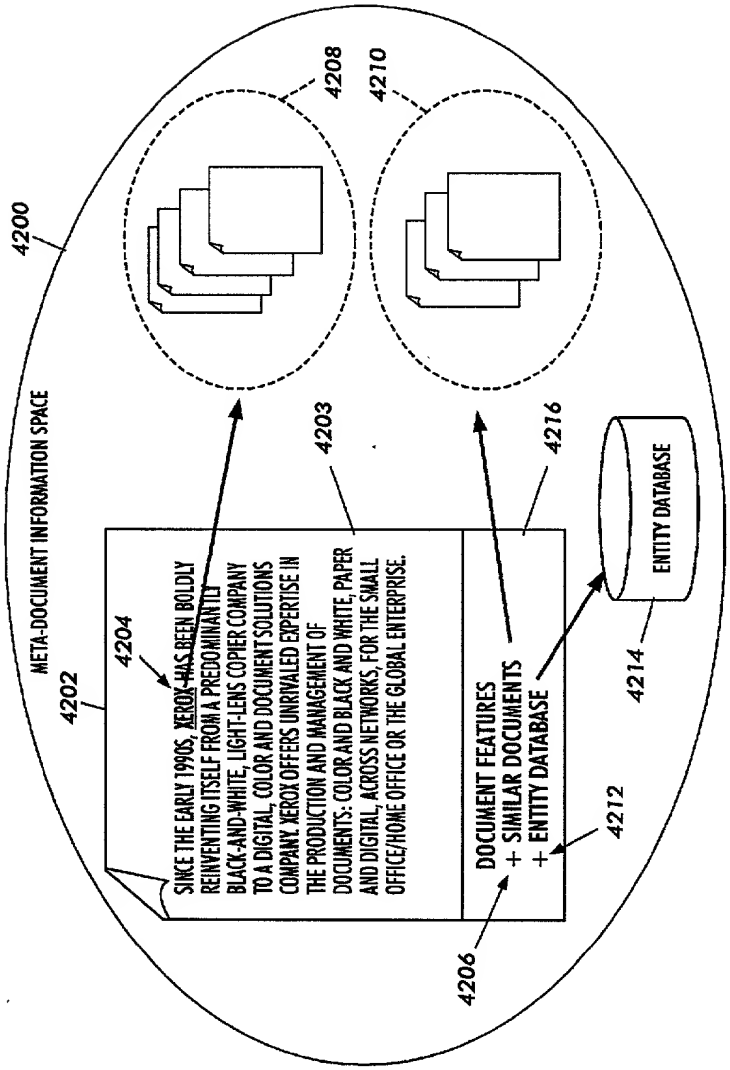
4200
META-DOCUMENT INFORMATION SPACE
4208
4210
4202
4204
SINCE THE EARLY 1990S, XEROX HAS BEEN BOLDLY REINVENTING ITSELF FROM A PREDOMINANTLY BLACK-AND-WHITE, LIGHT-LENS COPIER COMPANY TO A DIGITAL, COLOR AND DOCUMENT SOLUTIONS COMPANY. XEROX OFFERS UNRIVALED EXPERTISE IN THE PRODUCTION AND MANAGEMENT OF DOCUMENTS: COLOR AND BLACK AND WHITE, PAPER AND DIGITAL, ACROSS NETWORKS, FOR THE SMALL OFFICE/HOME OFFICE OR THE GLOBAL ENTERPRISE.
4203
4216
DOCUMENT FEATURES
+ SIMILAR DOCUMENTS
+ ENTITY DATABASE
4212
4206
ENTITY DATABASE
4214

FIG. 42

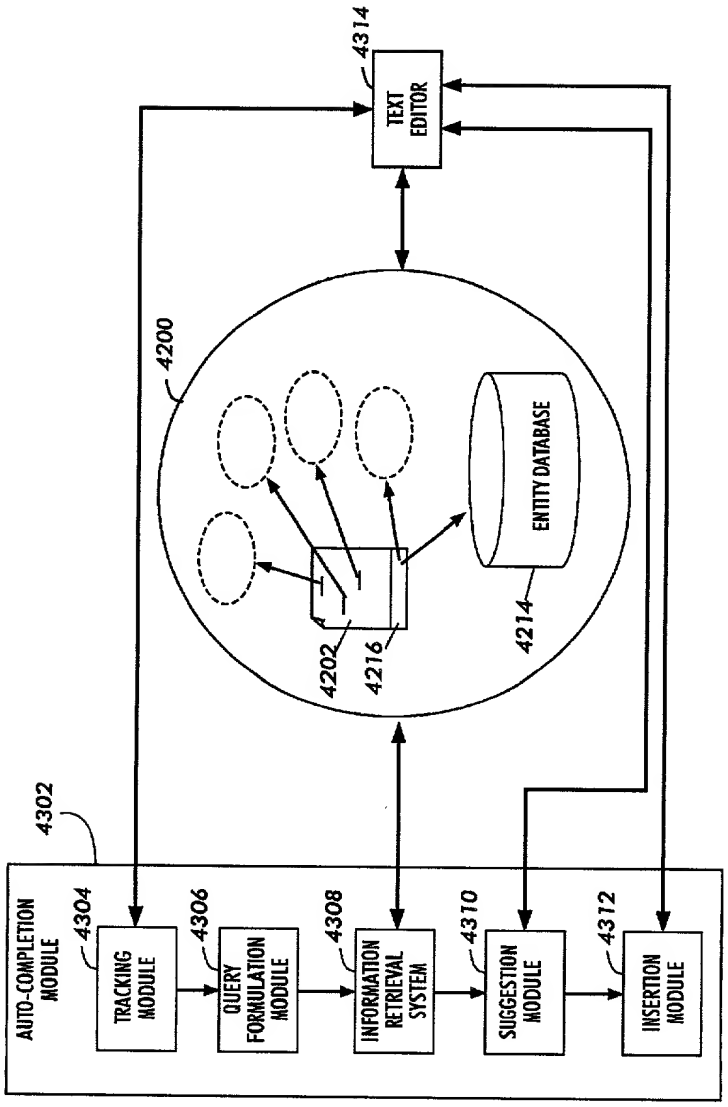
4314
TEXT EDITOR
4200
ENTITY DATABASE
4214
4202
4216
4302
AUTO-COMPLETION MODULE
4304
TRACKING MODULE
4306
QUERY FORMULATION MODULE
4308
INFORMATION RETRIEVAL SYSTEM
4310
SUGGESTION MODULE
4312
INSERTION MODULE

FIG. 43

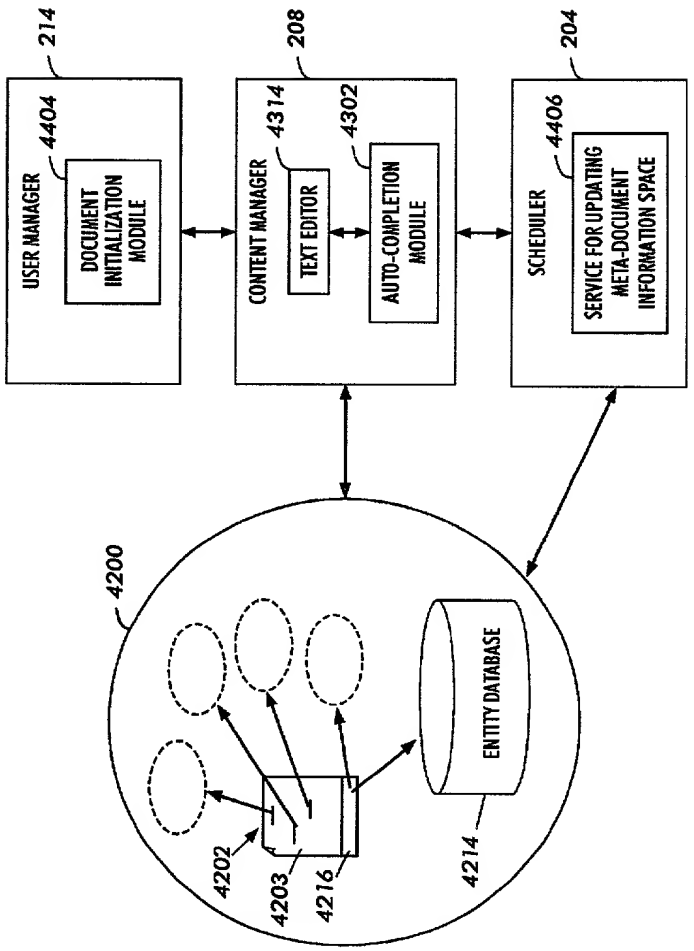
214
USER MANAGER
4404
DOCUMENT INITIALIZATION MODULE
208
CONTENT MANAGER
4314
TEXT EDITOR
4302
AUTO-COMPLETION MODULE
204
SCHEDULER
4406
SERVICE FOR UPDATING META-DOCUMENT INFORMATION SPACE
4200
ENTITY DATABASE
4202
4203
4216
4214

FIG. 44

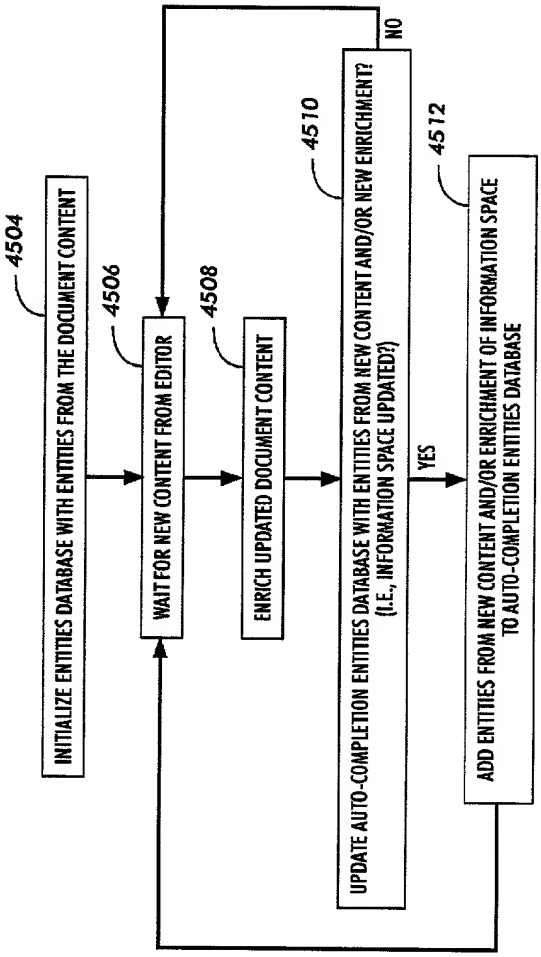
4504
INITIALIZE ENTITIES DATABASE WITH ENTITIES FROM THE DOCUMENT CONTENT
4506
WAIT FOR NEW CONTENT FROM EDITOR
4508
ENRICH UPDATED DOCUMENT CONTENT
4510
UPDATE AUTO-COMPLETION ENTITIES DATABASE WITH ENTITIES FROM NEW CONTENT AND/OR NEW ENRICHMENT? (I.E., INFORMATION SPACE UPDATED?)
NO
YES
4512
ADD ENTITIES FROM NEW CONTENT AND/OR ENRICHMENT OF INFORMATION SPACE TO AUTO-COMPLETION ENTITIES DATABASE

FIG. 45

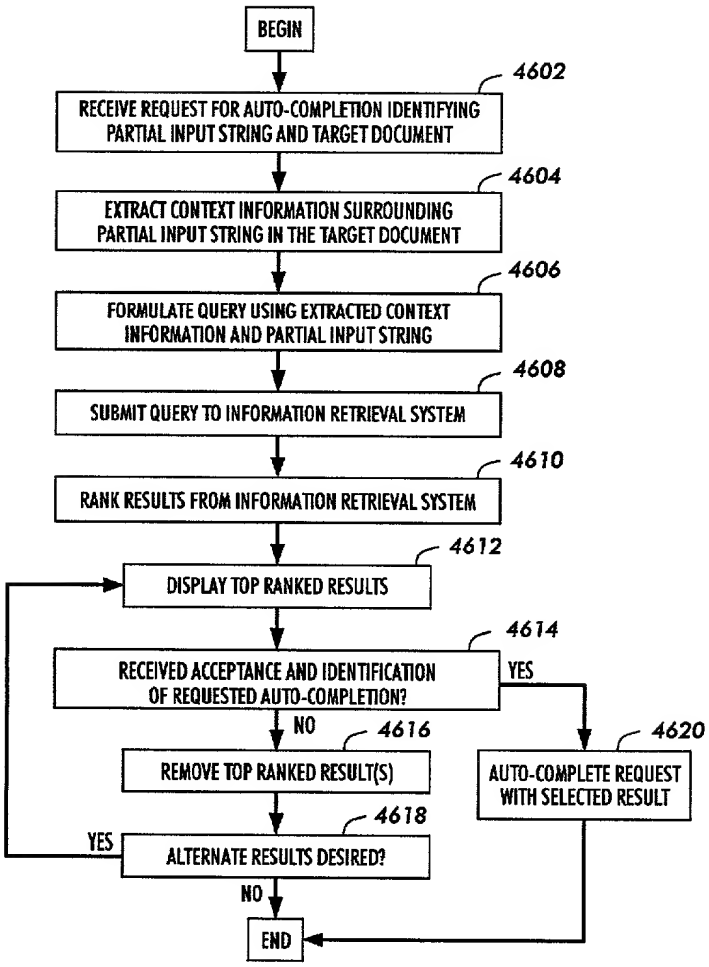
BEGIN
4602
RECEIVE REQUEST FOR AUTO-COMPLETION IDENTIFYING PARTIAL INPUT STRING AND TARGET DOCUMENT
4604
EXTRACT CONTEXT INFORMATION SURROUNDING PARTIAL INPUT STRING IN THE TARGET DOCUMENT
4606
FORMULATE QUERY USING EXTRACTED CONTEXT INFORMATION AND PARTIAL INPUT STRING
4608
SUBMIT QUERY TO INFORMATION RETRIEVAL SYSTEM
4610
RANK RESULTS FROM INFORMATION RETRIEVAL SYSTEM
4612
DISPLAY TOP RANKED RESULTS
4614
RECEIVED ACCEPTANCE AND IDENTIFICATION OF REQUESTED AUTO-COMPLETION?
YES
NO
4616
REMOVE TOP RANKED RESULT(S)
4618
ALTERNATE RESULTS DESIRED?
YES
NO
4620
AUTO-COMPLETE REQUEST WITH SELECTED RESULT
END

FIG. 46

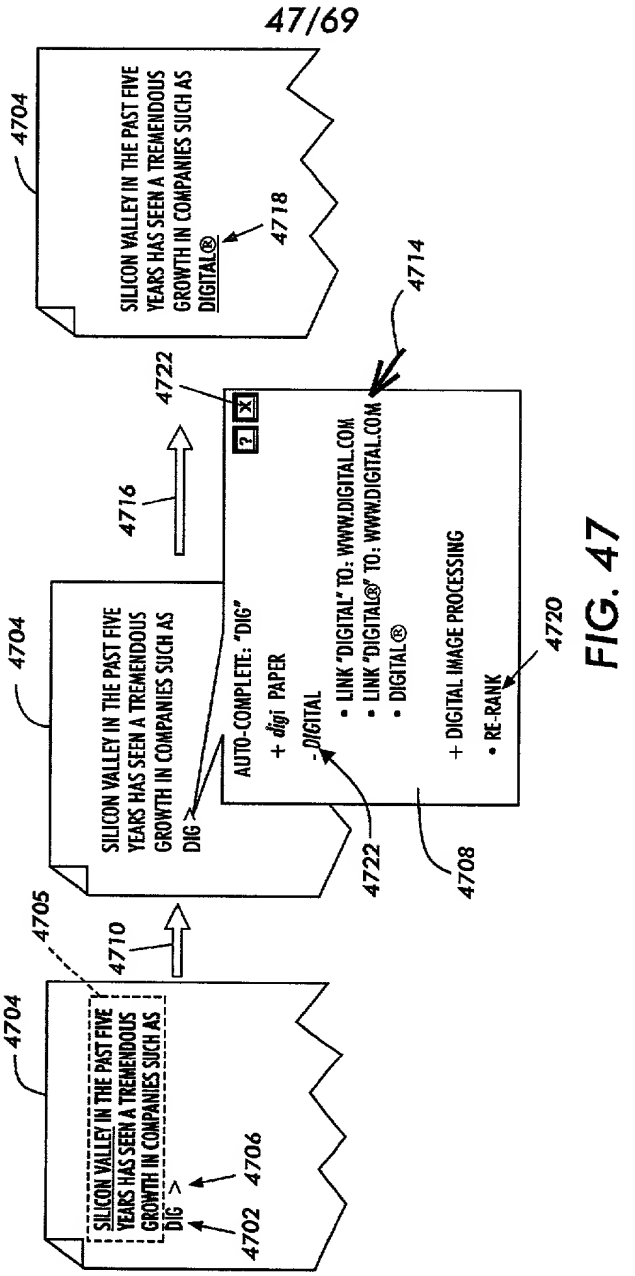
4704
SILICON VALLEY IN THE PAST FIVE YEARS HAS SEEN A TREMENDOUS GROWTH IN COMPANIES SUCH AS DIGITAL®
4718
4714
4722
4716
4704
SILICON VALLEY IN THE PAST FIVE YEARS HAS SEEN A TREMENDOUS GROWTH IN COMPANIES SUCH AS DIG
AUTO-COMPLETE: "DIG"
+ digi PAPER
- DIGITAL
• LINK "DIGITAL" TO: WWW.DIGITAL.COM
• LINK "DIGITAL®" TO: WWW.DIGITAL.COM
• DIGITAL®
+ DIGITAL IMAGE PROCESSING
• RE-RANK
4720
4722
4708
4705
4710
4704
SILICON VALLEY IN THE PAST FIVE YEARS HAS SEEN A TREMENDOUS GROWTH IN COMPANIES SUCH AS DIG
4706
4702

FIG. 47

(A) KEY	(B) ENTITY	(C) MARKUP	(D) FORMAT	(E) ORIGIN	(F) ORIGIN POSITION	(G) PART OF SPEECH	(H) CONTEXT
Sili	Silicon Valley	www.siliconvalley.com	Capitalize	TARGET DOCUMENT CONTENT	1	Noun	Science
Dig	Digital	www.digital.com	®	www.digital.com	434	Adjective	Science
Dig	DigiPaper	NONE	Bold	www.parc.xerox.com	23	Noun	Science
Dig	digress	NONE	NONE	www.nytimes.com	343	Verb	Politics
...	...	...	...	...	...	...	...

FIG. 48

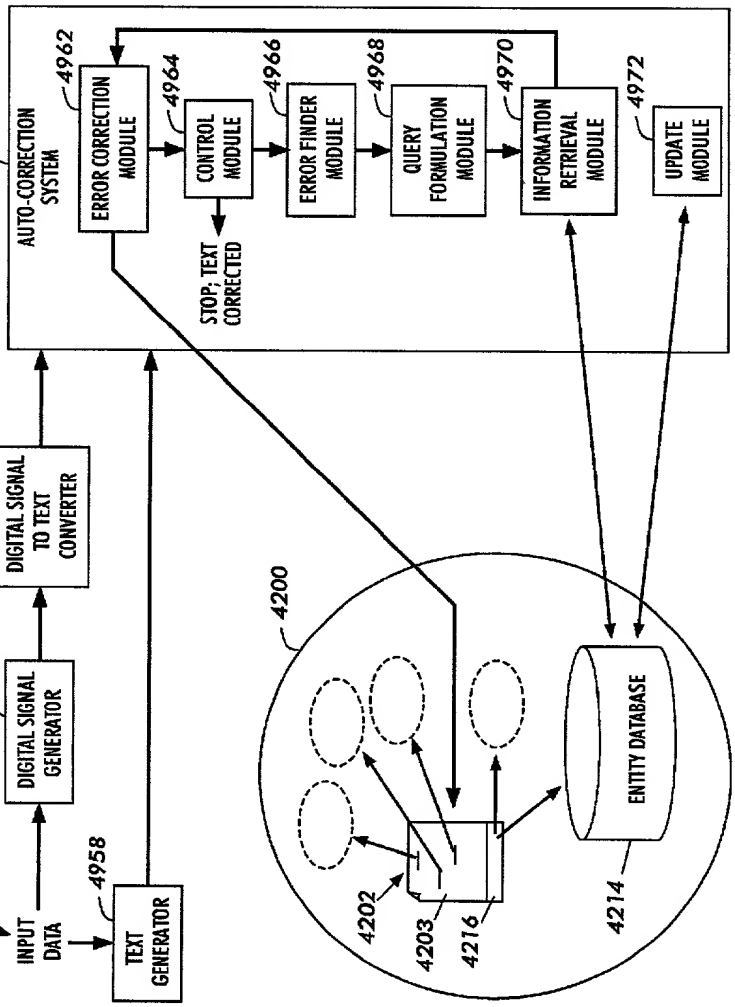
4960
AUTO-CORRECTION SYSTEM
4962
ERROR CORRECTION MODULE
4964
CONTROL MODULE
STOP; TEXT CORRECTED
4966
ERROR FINDER MODULE
4968
QUERY FORMULATION MODULE
4970
INFORMATION RETRIEVAL MODULE
4972
UPDATE MODULE
4956
DIGITAL SIGNAL TO TEXT CONVERTER
4954
DIGITAL SIGNAL GENERATOR
4952
INPUT DATA
4958
TEXT GENERATOR
4200
4202
4203
4216
4214
ENTITY DATABASE

FIG. 49

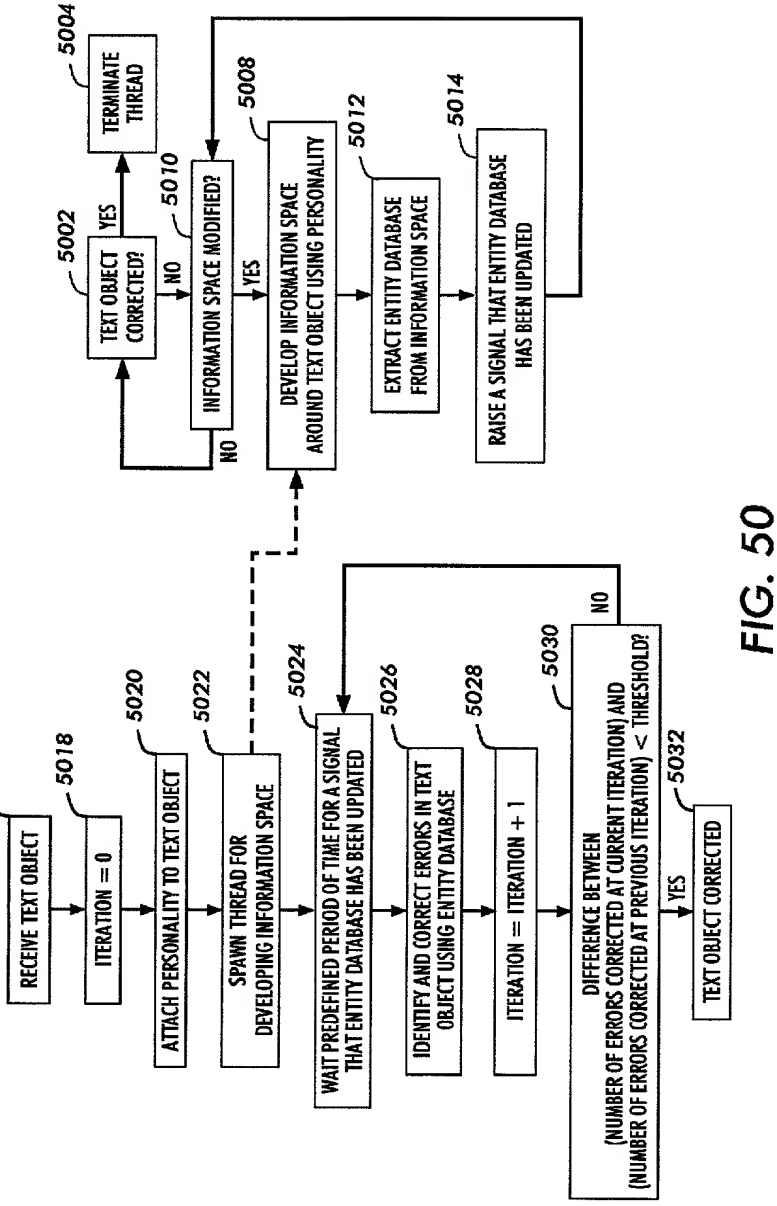
5000
RECEIVE TEXT OBJECT
5018
ITERATION = 0
5020
ATTACH PERSONALITY TO TEXT OBJECT
5022
SPAWN THREAD FOR DEVELOPING INFORMATION SPACE
5024
WAIT PREDEFINED PERIOD OF TIME FOR A SIGNAL THAT ENTITY DATABASE HAS BEEN UPDATED
5026
IDENTIFY AND CORRECT ERRORS IN TEXT OBJECT USING ENTITY DATABASE
5028
ITERATION = ITERATION + 1
5030
DIFFERENCE BETWEEN (NUMBER OF ERRORS CORRECTED AT CURRENT ITERATION) AND (NUMBER OF ERRORS CORRECTED AT PREVIOUS ITERATION) < THRESHOLD?
NO
YES
5032
TEXT OBJECT CORRECTED
5008
DEVELOP INFORMATION SPACE AROUND TEXT OBJECT USING PERSONALITY
5012
EXTRACT ENTITY DATABASE FROM INFORMATION SPACE
5014
RAISE A SIGNAL THAT ENTITY DATABASE HAS BEEN UPDATED
5010
INFORMATION SPACE MODIFIED?
YES
NO
5002
TEXT OBJECT CORRECTED?
YES
NO
5004
TERMINATE THREAD

FIG. 50

51/69

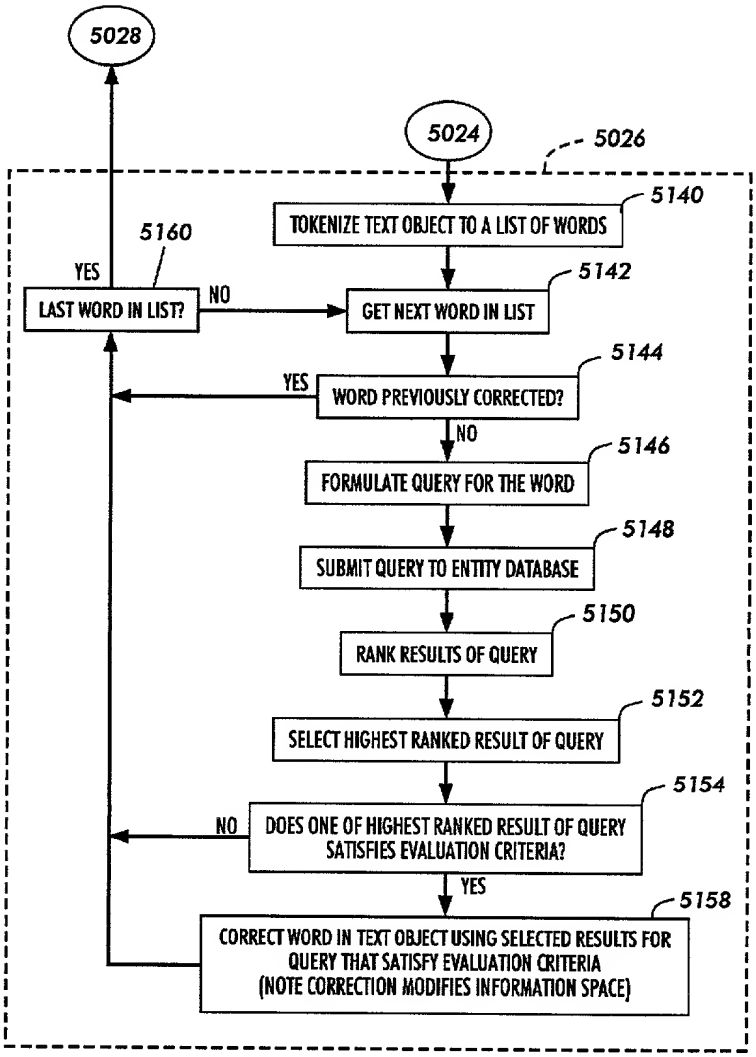
5028
5024
5026
5140
TOKENIZE TEXT OBJECT TO A LIST OF WORDS
5160
YES
LAST WORD IN LIST?
NO
5142
GET NEXT WORD IN LIST
5144
YES
WORD PREVIOUSLY CORRECTED?
NO
5146
FORMULATE QUERY FOR THE WORD
5148
SUBMIT QUERY TO ENTITY DATABASE
5150
RANK RESULTS OF QUERY
5152
SELECT HIGHEST RANKED RESULT OF QUERY
5154
NO
DOES ONE OF HIGHEST RANKED RESULT OF QUERY SATISFIES EVALUATION CRITERIA?
YES
5158
CORRECT WORD IN TEXT OBJECT USING SELECTED RESULTS FOR QUERY THAT SATISFY EVALUATION CRITERIA
(NOTE CORRECTION MODIFIES INFORMATION SPACE)

FIG. 51

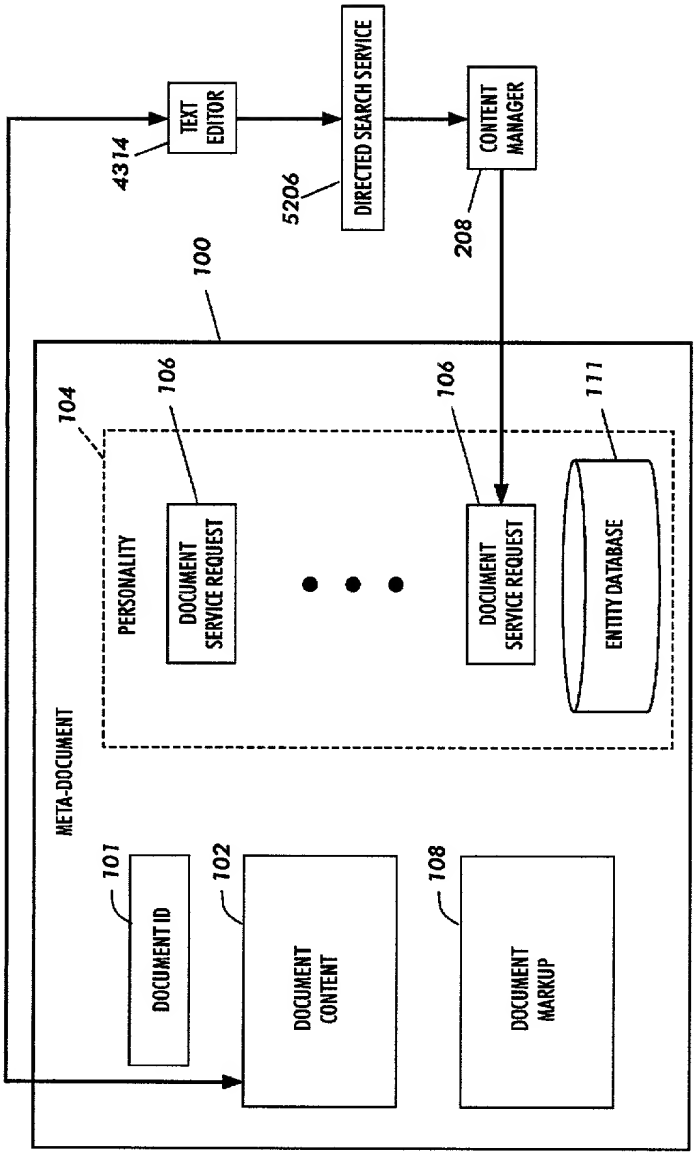
TEXT EDITOR
DIRECTED SEARCH SERVICE
CONTENT MANAGER
4314
5206
208
100
META-DOCUMENT
104
PERSONALITY
106
DOCUMENT SERVICE REQUEST
106
DOCUMENT SERVICE REQUEST
111
ENTITY DATABASE
101
DOCUMENT ID
102
DOCUMENT CONTENT
108
DOCUMENT MARKUP

FIG. 52

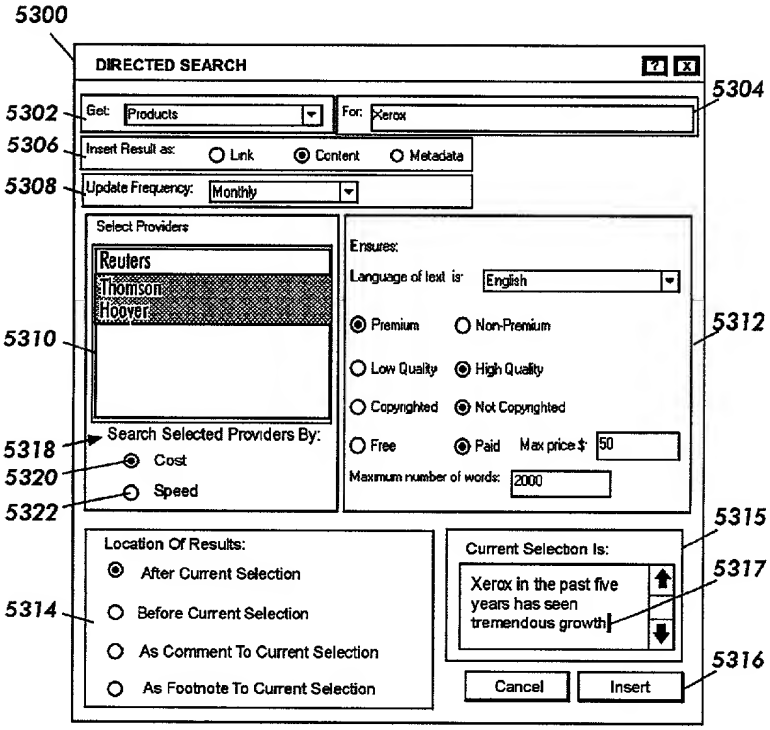
5300
DIRECTED SEARCH
5302
Get: Products
For: Xerox
5304
5306
Insert Result as: Link Content Metadata
5308
Update Frequency: Monthly
Select Providers
Reuters
Thomson
Hoover
5310
Ensures:
Language of text is: English
Premium Non-Premium
5312
Low Quality High Quality
Copyrighted Not Copyrighted
5318
Search Selected Providers By:
Free Paid Max price $: 50
5320
Cost
Maximum number of words: 2000
5322
Speed
5315
Location Of Results:
Current Selection Is:
After Current Selection
5317
5314
Before Current Selection
Xerox in the past five years has seen tremendous growth
As Comment To Current Selection
As Footnote To Current Selection
5316
Cancel
Insert

FIG. 53

5400

```
<service id='Products'
    refresh=30
    insert='Content'
    language='En'
    quality='Premium&High&Not-copyrighted&paid'
    maxprice='50'
    maxwords='2000'
    search='cost'>
  <entity>Xerox
    <offsets>13 18 456 461 710 715</offsets>
    <provider name='Thomson' id=18></provider>
    <provider name='Reuters' id=19></provider>
  </entity>
</service>
```

FIG. 54

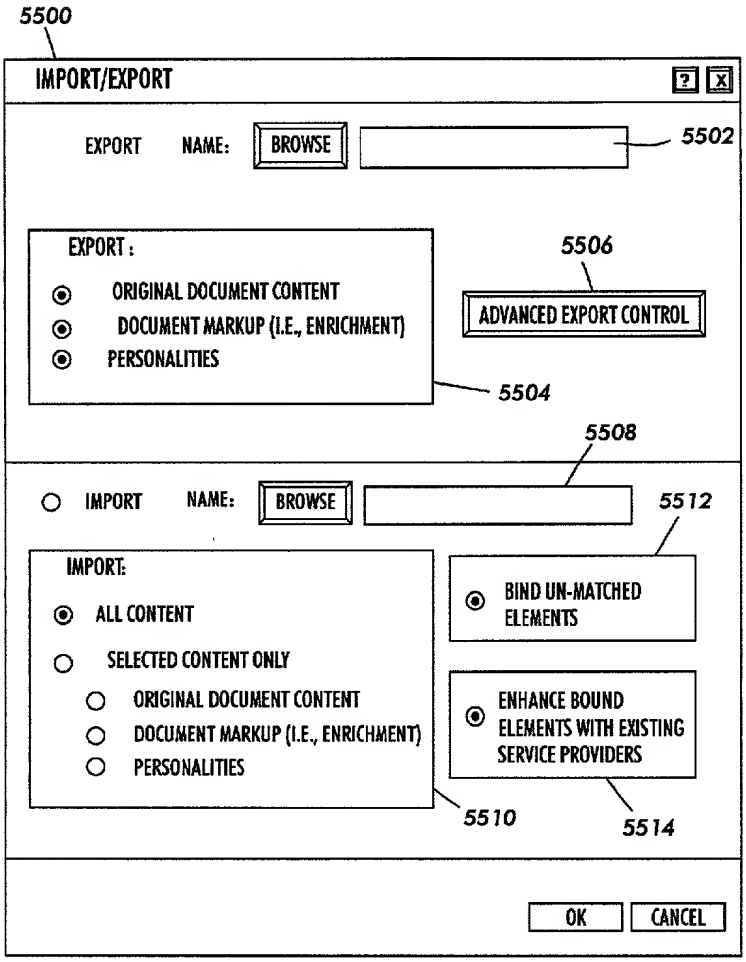
5500
IMPORT/EXPORT
EXPORT
NAME:
BROWSE
5502
EXPORT:
ORIGINAL DOCUMENT CONTENT
DOCUMENT MARKUP (I.E., ENRICHMENT)
PERSONALITIES
5504
5506
ADVANCED EXPORT CONTROL
5508
IMPORT
NAME:
BROWSE
5512
IMPORT:
ALL CONTENT
SELECTED CONTENT ONLY
ORIGINAL DOCUMENT CONTENT
DOCUMENT MARKUP (I.E., ENRICHMENT)
PERSONALITIES
5510
BIND UN-MATCHED ELEMENTS
ENHANCE BOUND ELEMENTS WITH EXISTING SERVICE PROVIDERS
5514
OK
CANCEL

FIG. 55

```
<!- SERVER ID>
<docsouls:serverid> DS v1.0 server name </docsouls:serverid>
    <!- PROPERTIES OF DOCUMENT>
    <docsouls:docid> xxx </docsouls:docid>
    <docsouls:name> xxx </docsouls:name>
<docsouls:owner> xxx </docsouls:owner>
...
<!-- ORIGINAL DOCUMENTS DOCUMENT CONTENT>
<docsouls:originaldoc format="HTML">toto.html</docsouls:originaldoc>
<docsouls:browsingdoc format="HTML">browse.html</docsouls:browsingdoc>

<!- LIST OF PERSONALITIES AND DOCUMENT SERVICE REQUESTS>
<docsouls:endowmentdescription>
    <!-- LIST OF PERSONALITIES>
    <docsouls:personality id = " " name="" etc.>
        <docsouls:service id="" name="" etc.>
            <docsouls:provider name="" etc.>
            <docsouls:provider name="" etc.>
        </docsouls:service>
    <docsouls:personality>
    ...
    <!-- LIST OF SERVICE REQUESTS>
    <docsouls:service id="" name="" etc.>
        <docsouls:provider name="" etc.>
        <docsouls:provider name="" etc.>
    </docsouls:service>
    ...
</docsouls:endowmentdescription>
```

FIG. 56

```
<!- DESCRIPTIONS OF PERSONALITIES>
<docsouls:personalitydescription>
    <docsouls:id> xx </docsouls:id>
      <!-- all the fields AND service categories>
      <docsouls:servicecategory id="" name = "">
</docsouls: personalitydescription>

<!- NEXT PERSONALITY DESCRIPTION>
...

<!- DESCRIPTION OF DOCUMENT SERVICE REQUESTS>
<docsouls:servicedescription>
    <docsouls:id> xx </docsouls:id>
      <!-- list all fields (e.g., "refresh period", category, etc.)>
      <!-- list all providers>
      <docsouls:provider id="" name = "">
      <!-- list all strategies>
      <!-- list status of services (original, frozen, mapped, etc.)>
      <docsouls:servicestatus status="mapped" originalid="" originalserver="">
      <!-- list ServerId if the service was not available on current server>
      <docsouls:serverid>id</docsouls:serverid>
</docsouls:servicedescription>

<!- NEXT SERVICE REQUESTS DESCRIPTION>
...
```

FIG. 57

```
<!- DESCRIPTION OF PROVIDERS>
<docsouls:providerdescription>
    <docsouls:id> xx </docsouls:id>
    <!-- all the fields e.g. key, URL, etc.>
</docsouls: providerdescription >

<!- DESCRIPTION OF DICTIONARIES>
<docsouls:dictionarydescription>
    <docsouls:id> xx </docsouls:id>
    <!-- all the fields, e.g. category>
</docsouls:dictionarydescription>

<!- DESCRIPTION OF STRATEGY>
<docsouls:strategydescription>
    <docsouls:id> xx </docsouls:id>
    <!-all the fields>
    <!-all the dictionaries>
</docsouls:strategydescription>

<!- RESULTS OF GLOBAL SERVICES >
<docsouls:globalservice id="Summary">
    <docsouls:link> xxx </docsouls:link>
    <docsouls:content> xxx </docsouls:content>
</docsouls:globalservice>

<!- OTHER RESULTS (E.G. KEYWORDS, CATEGORIES, ETC.>
```

FIG. 58A

```
<!- (OPTIONAL) DESCRIPTION OF ENRICHMENT PERFORMED AT OTHER DOCUMENT ENRICHMENT
SYSTEMS>
<docsouls:serverid>olderServerId</docsouls:serverid>
<docsouls:personalitydescription>...</docsouls: personalitydescription>
<docsouls:servicedescription>...</docsouls:servicedescription>

<!- (OPTIONAL) LIST OF RESULTS OF THE LAST ENRICHMENT >
    <!- IDENTIFY DOCUMENT SECTION IN CASE OF MULTI-PART DOC>
    <docsouls:docpart>main_frame.html</docsouls:docpart>
    <!- OFFSET WHEN ENRICHMENT APPLIES TO A SECTION>
    <docsouls:offset> 0 200 </docsouls:offset>
    <!- STATUS OF SERVICE APPLIED ("ACTIVE" OR "FROZEN") >
    <docsouls:service id="StockQuote" status="active">
        <!-- ENTITY>
        <docsouls:entity>Xerox
            <!-- NORMALIZED FORM >
            <docsouls:normform>Xerox</docsouls:normform>
            <!-- LIST OF OFFSETS WHERE ENTITY IS FOUND IN TEXT RELATIVE TO SECTION >
            <docsouls:offsets start=35 end=40 />
            <docsouls:offsets start=232 end=237 />
            <!-- DICTIONARIES >
            <docsouls:dict> xxx </docsouls:dict>
            <!-- STRATEGIES >
            <docsouls:strat> xxx </docsouls:strat>
            <!-- LINK INFORMATION >
            <docsouls:link> xxx </docsouls:link>
            <!-- CONTENT INFORMATION >
            <docsouls:content> xxx </docsouls:content>
        </docsouls:entity>
        <!-- NEXT ENTITY>
        ...
    </docsouls:service>
```

FIG. 58B

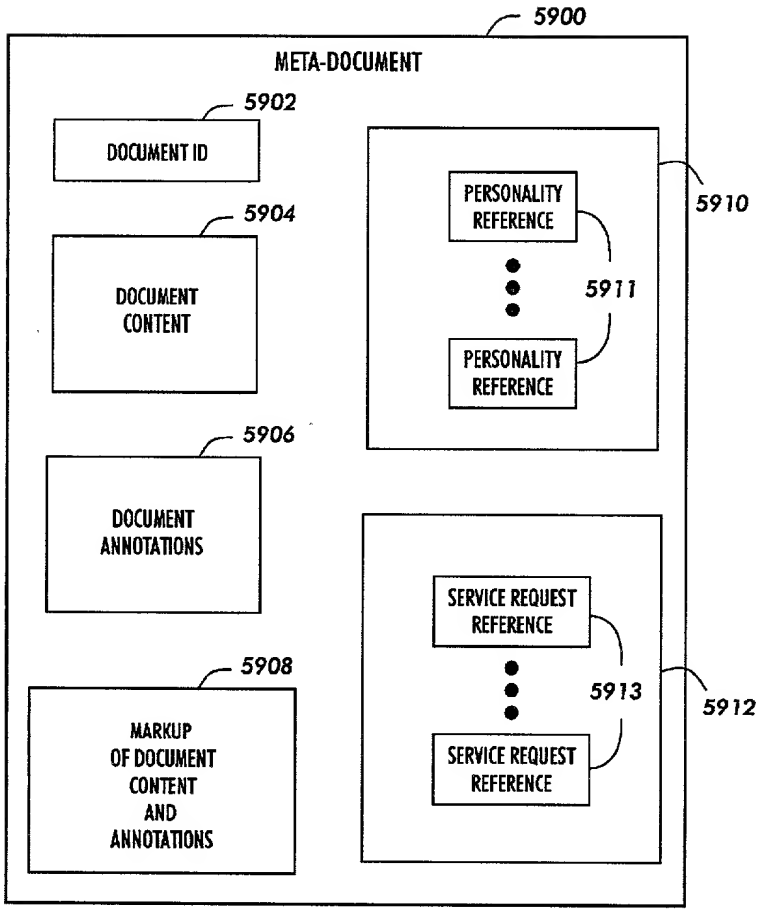
5900
META-DOCUMENT
5902
DOCUMENT ID
5904
DOCUMENT CONTENT
5906
DOCUMENT ANNOTATIONS
5908
MARKUP OF DOCUMENT CONTENT AND ANNOTATIONS
PERSONALITY REFERENCE
5911
PERSONALITY REFERENCE
5910
SERVICE REQUEST REFERENCE
5913
SERVICE REQUEST REFERENCE
5912

FIG. 59

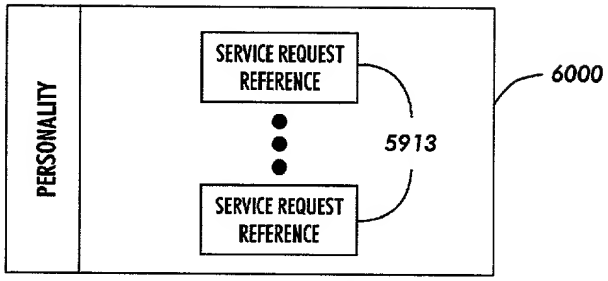
PERSONALITY
SERVICE REQUEST REFERENCE
SERVICE REQUEST REFERENCE
5913
6000

FIG. 60

62/69

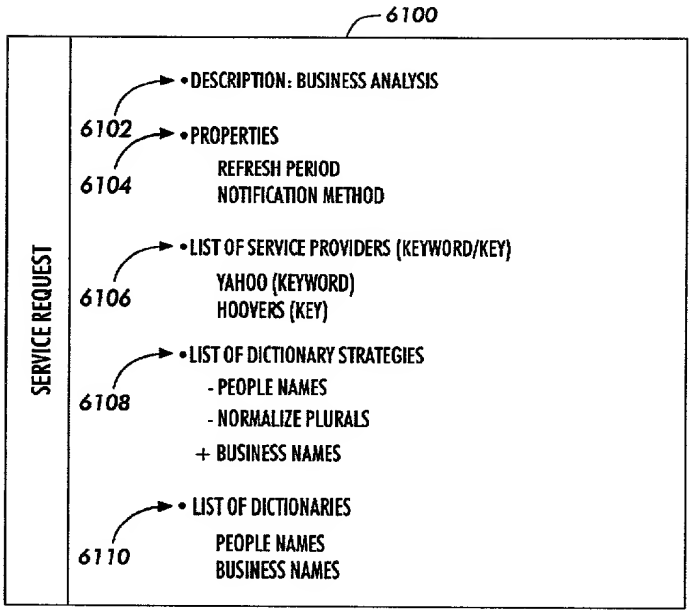
6100
SERVICE REQUEST
•DESCRIPTION: BUSINESS ANALYSIS
6102
•PROPERTIES
6104
REFRESH PERIOD
NOTIFICATION METHOD
•LIST OF SERVICE PROVIDERS (KEYWORD/KEY)
6106
YAHOO (KEYWORD)
HOOVERS (KEY)
•LIST OF DICTIONARY STRATEGIES
6108
- PEOPLE NAMES
- NORMALIZE PLURALS
+ BUSINESS NAMES
• LIST OF DICTIONARIES
6110
PEOPLE NAMES
BUSINESS NAMES

FIG. 61

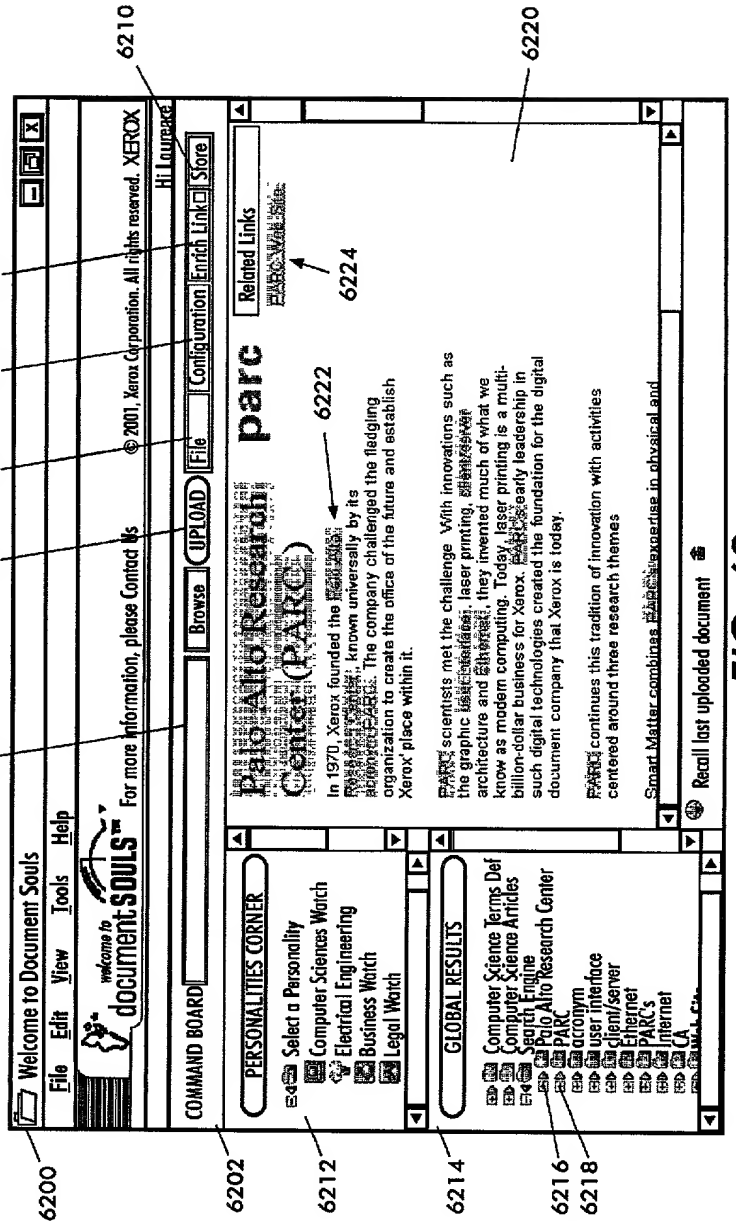
6200
6202
6203
6204
6205
6206
6208
6210
6212
6214
6216
6218
6220
6222
6224
Welcome to Document Souls
File Edit View Tools Help
welcome to documentSOULS™
For more information, please Contact Us
© 2001, Xerox Corporation. All rights reserved. XEROX
Hi Lourdes
COMMAND BOARD
Browse
UPLOAD
File
Configuration
Enrich Link
Store
PERSONALITIES CORNER
Select a Personality
Computer Sciences Watch
Electrical Engineering
Business Watch
Legal Watch
GLOBAL RESULTS
Computer Science Terms Def
Computer Science Articles
Search Engine
Palo Alto Research Center
PARC
acronym
user interface
client/server
Ethernet
PARC's
Internet
CA
parc
Related Links
In 1970, Xerox founded the
known universally by its
. The company challenged the fledgling
organization to create the office of the future and establish
Xerox' place within it.
scientists met the challenge. With innovations such as
the graphic
, laser printing,
architecture and
, they invented much of what we
know as modern computing. Today, laser printing is a multi-
billion-dollar business for Xerox.
such digital technologies created the foundation for the digital
document company that Xerox is today.
continues this tradition of innovation with activities
centered around three research themes
Smart Matter combines
Recall last uploaded document

FIG. 62

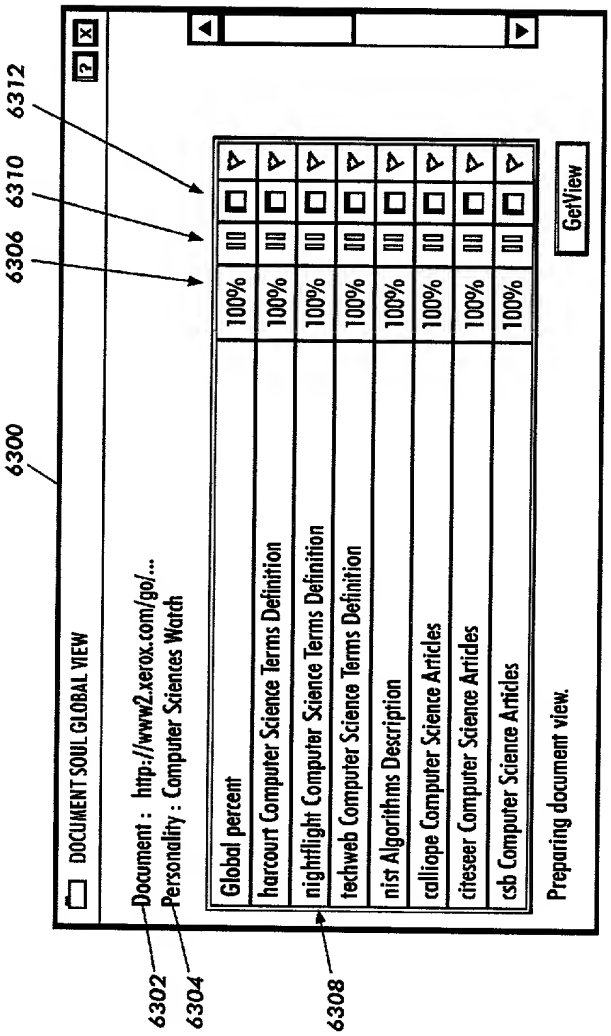
6300
6302
6304
6306
6308
6310
6312
DOCUMENT SOUL GLOBAL VIEW
Document : http://www2.xerox.com/go/...
Personality : Computer Sciences Watch
Global percent
harcourt Computer Science Terms Definition
nightflight Computer Science Terms Definition
techweb Computer Science Terms Definition
nist Algorithms Description
calliope Computer Science Articles
citeseer Computer Science Articles
csb Computer Science Articles
100%
GetView
Preparing document view.

FIG. 63

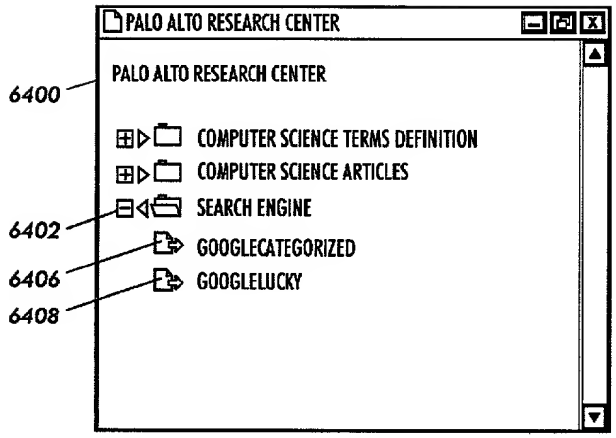
PALO ALTO RESEARCH CENTER
PALO ALTO RESEARCH CENTER
6400
COMPUTER SCIENCE TERMS DEFINITION
COMPUTER SCIENCE ARTICLES
SEARCH ENGINE
6402
GOOGLECATEGORIZED
6406
GOOGLELUCKY
6408

FIG. 64

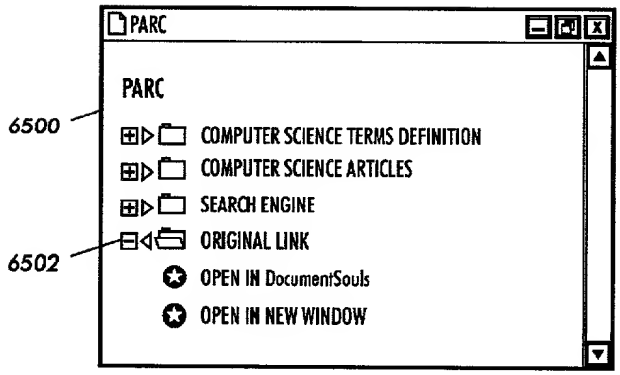
PARC
PARC
6500
COMPUTER SCIENCE TERMS DEFINITION
COMPUTER SCIENCE ARTICLES
SEARCH ENGINE
6502
ORIGINAL LINK
OPEN IN DocumentSouls
OPEN IN NEW WINDOW

FIG. 65

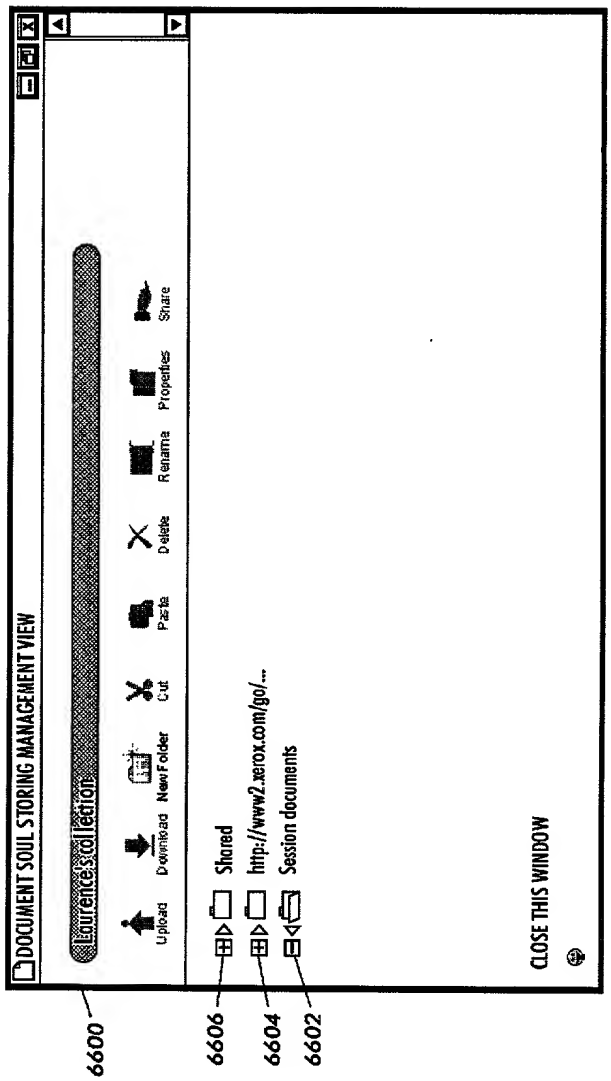
DOCUMENT SOUL STORING MANAGEMENT VIEW
Laurence's collections
Upload
Download
New Folder
Cut
Paste
Delete
Rename
Properties
Share
Shared
http://www2.xerox.com/go/...
Session documents
CLOSE THIS WINDOW
6600
6606
6604
6602

FIG. 66

68/69

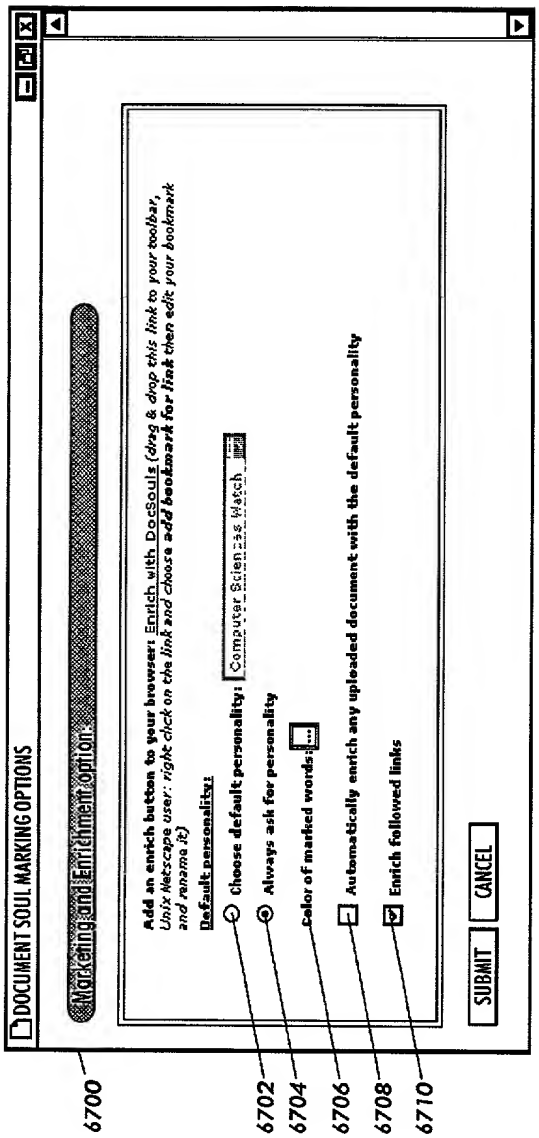
DOCUMENT SOUL MARKING OPTIONS
Marketing and Enrichment options
Add an enrich button to your browser: Enrich with DocSouls (drag & drop this link to your toolbar, Unix Netscape user: right click on the link and choose add bookmark for link then edit your bookmark and rename it)
Default personality:
Choose default personality:
Computer Sciences Watch
Always ask for personality
Color of marked words:
Automatically enrich any uploaded document with the default personality
Enrich followed links
SUBMIT
CANCEL
6700
6702
6704
6706
6708
6710

FIG. 67

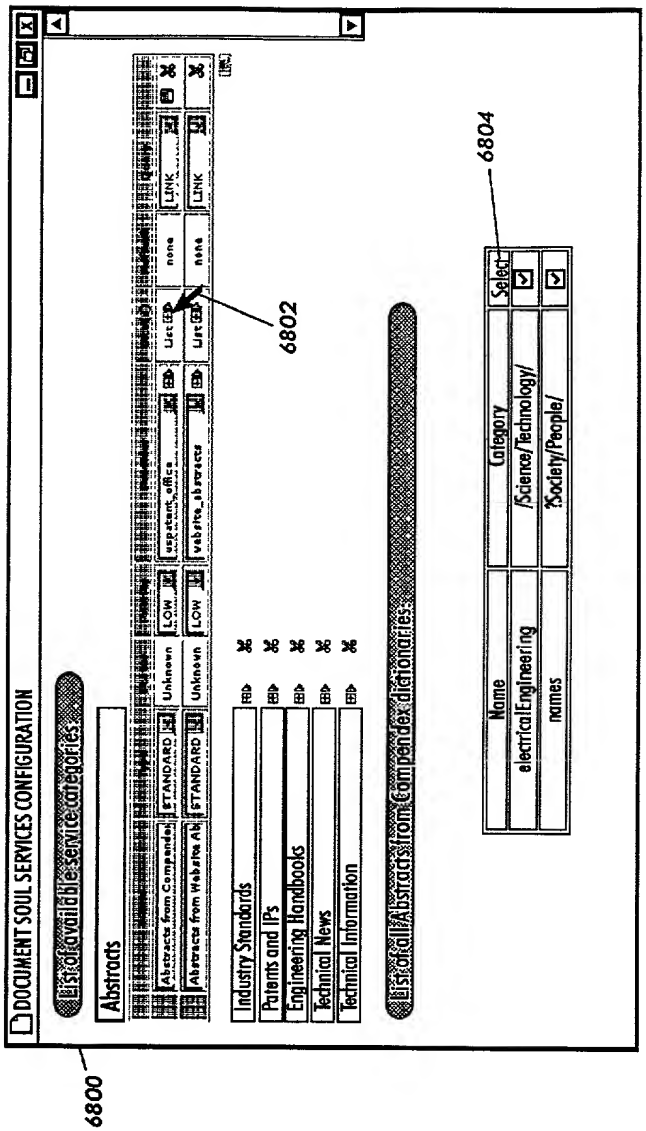
DOCUMENT SOUL SERVICES CONFIGURATION
List of available service categories:
Abstracts
Industry Standards
Patents and IPs
Engineering Handbooks
Technical News
Technical Information
List of all Abstracts from compendex dictionaries:
Name
electricalEngineering
names
Category
/Science/Technology/
/Society/People/
Select
6800
6802
6804

FIG. 68